

UNIVERSIDADE FEDERAL DO PARANÁ

LORENA CARVALHO DA PAZ

MODELO MATEMÁTICO PARA TEMPO DE TRANSFORMAÇÃO DE
PRIORIDADES NO DESLOCAMENTO INTRA-HOSPITALAR

CURITIBA

2021

LORENA CARVALHO DA PAZ

MODELO MATEMÁTICO PARA TEMPO DE TRANSFORMAÇÃO DE
PRIORIDADES NO DESLOCAMENTO INTRA-HOSPITALAR

Dissertação de Mestrado apresentada como requisito à obtenção do grau de Mestre, no Programa de Pós-Graduação em Engenharia de Produção, área de concentração Pesquisa Operacional, Setor de Tecnologia, Universidade Federal do Paraná.

Orientador: Prof. Dr. José Eduardo Pécora Júnior
Coorientador: Prof. Dr. Gustavo Valentim Loch

CURITIBA

2021

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

P348m Paz, Lorena Carvalho da
Modelo matemático para tempo de transformação de prioridades
no deslocamento intra-hospitalar [recurso eletrônico] / Lorena
Carvalho da Paz – Curitiba, 2021.

Dissertação - Universidade Federal do Paraná, Setor de
Tecnologia, Programa de Pós-graduação em Engenharia de
Produção.

Orientador: Prof. Dr. José Eduardo Pécora Júnior

Coorientador: Prof. Dr. Gustavo Valentim Loch

1. Transporte (Paciente hospitalar) 2. Transporte (Prioridade). I.
Universidade Federal do Paraná. II. Pécora Júnior, José Eduardo. III.
Loch, Gustavo Valentim. IV. Título.

CDD: 519.72

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585



MINISTÉRIO DA EDUCAÇÃO
SETOR DE TECNOLOGIA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA DE
PRODUÇÃO - 40001016070P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE PRODUÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **LORENA CARVALHO DA PAZ** intitulada: **MODELO MATEMÁTICO PARA TEMPO DE TRANSFORMAÇÃO DE PRIORIDADES NO DESLOCAMENTO INTRA-HOSPITALAR**, sob orientação do Prof. Dr. JOSÉ EDUARDO PÉCORA JUNIOR, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 26 de Maio de 2021.

Assinatura Eletrônica

27/05/2021 12:53:32.0

JOSÉ EDUARDO PÉCORA JUNIOR
Presidente da Banca Examinadora

Assinatura Eletrônica

27/05/2021 09:41:09.0

CASSIUS TADEU SCARPIN
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

27/05/2021 09:40:23.0

THIAGO ANDRÉ GUIMARÃES
Avaliador Externo (INSTITUTO FEDERAL DE EDUC., CIÊNCIA E
TECNOLOGIA DO PARANÁ)

Assinatura Eletrônica

27/05/2021 11:54:58.0

LEANDRO MAGATÃO
Avaliador Externo (UNIVERSIDADE TECNOLÓGICA FEDERAL DO
PARANÁ)

e insira o código 93964

AGRADECIMENTOS

Agradeço a Deus pela vida, saúde, sustento, cuidado e pelas oportunidades vividas;

a minha família Pedro, Cristina, Camila e Eliane por todo o amor e apoio em todos os momentos. Nunca mediram esforços para me proporcionar a melhor qualidade nos estudos, sustento em toda a vida. Também agradeço ao meu noivo Johnathan e meus amigos pela parceria e por me apoiarem na rotina vivida esses últimos 2 anos;

ao Professor Eduardo Pécora que não mediu esforços para me ajudar a trilhar o caminho da pesquisa operacional. Todo o apoio, torcida, tempo dedicado e principalmente pelo desenvolvimento profissional e acadêmico que vivi nesse período;

ao Professor Gustavo pela dedicação e auxílio. Sempre disponível para ajudar no desenvolvimento mais específico da dissertação e olhar crítico;

ao Grupo de Tecnologia Aplicada à Otimização (GTAO). Esse grupo de pesquisa proporciona um ambiente de troca de experiências e apoio que foram fundamentais para meu desenvolvimento e crescimento nesse período de mestrado;

ao meu amigo Vinícius Ton por todo o apoio na parte computacional do trabalho. Pela disposição e auxílio no uso da ferramenta de design de experimentos *irace*. Também auxiliou cedendo e explicando seu artigo que serviu de base para o trabalho aqui desenvolvido. Não mediu esforços para ajudar, não importava o dia da semana ou horário;

a Nathália Ortiz pelo apoio também cedendo e explicando sobre sua tese de doutorado e artigo que serviram de base para o trabalho aqui desenvolvido;

ao Guilherme do GTAO que foi fundamental no início do desenvolvimento do código auxiliando na leitura das instâncias;

a CAPES pelo suporte financeiro garantindo tranquilidade para o desenvolvimento da pesquisa, bem como minha dedicação 100% para esse fim.

RESUMO

O transporte intra-hospitalar de pacientes realizado em macas é um processo importante e que requer atenção pois o tempo desse deslocamento pode influenciar no agravamento do estado de saúde dos pacientes. Existe um número de funcionários (maqueiros) para realizar o deslocamento e cada paciente tem um nível de prioridade para ser deslocado. Cada nível de prioridade tem um tempo máximo em que o paciente deve ser deslocado e quando esse tempo é ultrapassado o paciente recebe uma prioridade mais urgente. Esse tempo se trata do tempo de transformação de prioridades. Baseado na solicitação de um hospital canadense e em um artigo inicial do tema, foi realizada uma expansão do modelo para, após esse tempo limite, o paciente receber um nível de prioridade mais urgente para ser atendido mais rapidamente. O problema abordado será o tempo de espera para deslocamento e foi resolvido como sendo um sistema clássico de máquinas paralelas (*parallel machine*). Apesar do tema ter algumas abordagens, nenhuma toma como base a transformação de níveis de prioridade na fila de atendimento. Portanto, neste trabalho foi proposto um modelo de programação linear inteira mista para a transformação de prioridades no deslocamento intra-hospitalar utilizando as instâncias propostas de um artigo inicial sobre o tema. Foi realizada a comparação entre 3 tipos de testes (do modelo original, modelo com transformação de prioridades e o modelo desenvolvido nesse trabalho com transformação e melhoria de tempo) e para análise dos resultados vários aspectos foram levados em consideração: maior tempo de término entre as atividades, função objetivo, tempo de atraso médio por prioridade, tempo de resposta por prioridade, tempo ocioso dos maqueiros e atraso por prioridade. Sabe-se que o modelo aqui desenvolvido apresentou melhoras em alguns aspectos e a maior contribuição do estudo foi a sugestão de modificação dos tempos de prioridade, principalmente na prioridade 4 que é a mais urgente, reduzindo assim os atrasos recorrentes. O estudo desses tempos é um dos pontos fundamentais para a tomada de decisão do atendimento do hospital em questão, bem como a busca de um deslocamento sem demoras e atrasos, respeitando a necessidade hospitalar de cada paciente.

Palavras-Chave: transporte de paciente, máquinas paralelas, prioridade.

ABSTRACT

Intra-hospital patient transport on stretchers is an important process that requires attention because the time of this displacement can influence the worsening of the patients' health status. There are a number of employees (stretcher bearers) to perform the displacement and each patient has a priority level to be moved. Each priority level has a maximum time in which the patient must be moved, and when this time is exceeded, the patient receives a more urgent priority. This time is the priority transformation time. Based on a request from a Canadian hospital and an early article on the subject, an expansion of the model was made so that after this time limit, the patient would receive a more urgent priority level to be seen more quickly. The problem addressed will be the displacement waiting time and was solved as a classic parallel machine system. Although the subject has some approaches, none of them is based on the transformation of priority levels in the queue. Therefore, in this work, a mixed integer linear programming model was proposed for the transformation of priorities in intra-hospital displacement using the proposed instances from an initial paper on the subject. A comparison between 3 types of tests (the original model, the model with priority transformation and the model developed in this work with transformation and time improvement) was performed and to analyze the results several aspects were taken into consideration: longest completion time between activities, objective function, average delay time per priority, response time per priority, idle time of waiters and delay per priority. It is known that the model developed here presented improvements in some aspects and the greatest contribution of the study was the suggestion to modify the priority times, especially in priority 4 which is the most urgent, thus reducing recurring delays. The study of these times is one of the fundamental points for decision making in the care of the hospital in question, as well as the search for a displacement without delays and delays, respecting the hospital needs of each patient.

Keywords: patient transportation, parallel machine, priority.

LISTA DE FIGURAS

FIGURA 1 – EXEMPLO DE UTILIZAÇÃO DO GRÁFICO DE GANTT (SOFTWARE).....	23
FIGURA 2 – EXEMPLO DE UTILIZAÇÃO DO GRÁFICO DE GANTT COM 8 MÁQUINAS E 20 JOBS.....	24
FIGURA 3 - RELAÇÃO ENTRE AS CARACTERÍSTICAS DE UM PROBLEMA DE PRODUÇÃO E UM PROBLEMA INTRA-HOSPITALAR.....	36
FIGURA 4 - FLUXOGRAMA DE FUNCIONAMENTO DO MODELO MATEMÁTICO DESENVOLVIDO.....	44
FIGURA 5 - MATRIZ COM TEMPO DE DESLOCAMENTO ENTRE OS LOCAIS DOS PACIENTES (MINUTOS).....	47
FIGURA 6 - POSIÇÃO DOS PACIENTES NO VETOR JOBS PARA CONSIDERAR COM $T = 0$	48
FIGURA 7 - SEQUENCIAMENTO GERADO PELO MODELO COM RELÓGIO = 0.....	48
FIGURA 8 - POSIÇÃO DOS PACIENTES NO VETOR JOBS PARA CONSIDERAR COM $T = 4$	49
FIGURA 9 - TABELA DE TRADUÇÃO PARA SEQUENCIAMENTO COM RELÓGIO = 4.....	49
FIGURA 10 - SEQUENCIAMENTO GERADO PELO MODELO COM RELÓGIO = 4.....	50
FIGURA 11 - POSIÇÃO DOS PACIENTES NO VETOR JOBS PARA CONSIDERAR COM $T = 7$	50
FIGURA 12 - TABELA DE TRADUÇÃO PARA SEQUENCIAMENTO COM RELÓGIO = 7.....	50
FIGURA 13 - SEQUENCIAMENTO GERADO PELO MODELO COM RELÓGIO = 7.....	51

FIGURA 14 - POSIÇÃO DOS PACIENTES NO VETOR JOBS PARA CONSIDERAR COM $T = 12$	51
FIGURA 15 - TABELA DE TRADUÇÃO PARA SEQUENCIAMENTO COM RELÓGIO = 12.....	51
FIGURA 16 - SEQUENCIAMENTO GERADO PELO MODELO COM RELÓGIO = 12.....	52
FIGURA 17 - GRÁFICO PARA ANÁLISE DE SENSIBILIDADE DO TEMPO DE TRANSFORMAÇÃO DA PRIORIDADE 4.....	69

LISTA DE TABELAS

TABELA 1 – TEMPOS POR NÍVEL DE PRIORIDADE.....	46
TABELA 2 - NÍVEIS DE PRIORIDADE POR COR.....	46
TABELA 3 - DADOS DE ENTRADA DO SISTEMA.....	47
TABELA 4 - RESUMO DO USO DO IRACE.....	57
TABELA 5 - TESTES REALIZADOS.....	57
TABELA 6 - MÉDIA DE MAKESPAN EM SEGUNDOS.....	58
TABELA 7 - FUNÇÃO OBJETIVO EM SEGUNDOS.....	59
TABELA 8 - ANÁLISE DO TEMPO DE RESPOSTA MÉDIO DE P1 (MINUTOS).....	61
TABELA 9 - ANÁLISE DO TEMPO DE RESPOSTA MÉDIO DE P2 (MINUTOS).....	62
TABELA 10 - ANÁLISE DO TEMPO DE RESPOSTA MÉDIO DE P3 (MINUTOS).....	62
TABELA 11 - ANÁLISE DO TEMPO DE RESPOSTA MÉDIO DE P4 (MINUTOS).....	63
TABELA 12 - PORCENTAGEM DE TEMPO OCIOSO DOS MAQUEIROS.....	64
TABELA 13 - PORCENTAGEM DE PACIENTES ATRASADOS NAS PRIORIDADES 1 E 2 (%)	66
TABELA 14 - PORCENTAGEM DE PACIENTES ATRASADOS NAS PRIORIDADES 3 E 4 (%)	67
TABELA 15 - ATRASO REAL DA PRIORIDADE 4 EM MINUTOS.....	68

LISTA DE ABREVIATURAS E SIGLAS

CTAS	<i>Canadian Emergency Department Triage and Acuity Scale</i>
DARP	<i>Dial-a-Ride Problem</i>
DE	Departamento de Emergência
EMS	Serviço Médico de Emergência
FO	Função Objetivo
HNSG	Hospital Nossa Senhora das Graças
MPS	Plano-Mestre de Produção
PLIM	Programação Linear Inteira Mista
PS	Pronto-Socorro
SOSPS-PS	Aplicativo de Atendimento de Pronto-Socorro
UTI	Unidade de Terapia Intensiva

SUMÁRIO

1. INTRODUÇÃO	15
1.1 OBJETIVOS	17
1.1.1 Objetivo Geral	17
1.1.2 Objetivos Específicos	17
1.2 JUSTIFICATIVA DO TRABALHO	17
1.3 ESTRUTURA DO TRABALHO	18
2. DESCRIÇÃO DO PROBLEMA	20
2.1 PROBLEMAS DE PRODUÇÃO	20
2.2 MÁQUINAS PARALELAS	21
3. REVISÃO DA LITERATURA	25
3.1 AGENDAMENTO DE TAREFAS (CIRURGIAS, ATENDIMENTOS)	26
3.2 <i>DIAL-A-RIDE PROBLEM</i>	27
3.3 FILA DE PRIORIDADE	28
3.4 FLUXO DE PACIENTES	30
3.5 TEMPO DE ESPERA	31
3.6 TRANSPORTE DE PACIENTES	32
4. O PROBLEMA DE TRANSPORTE DE PACIENTES DENTRO DE HOSPITAIS	35
4.1 DEFINIÇÃO DO PROBLEMA	35
4.2 MODELO MATEMÁTICO DE PROGRAMAÇÃO LINEAR INTEIRA MISTA	37
4.2.1 Formulação do Modelo Matemático	38
4.2.2 Diferenças entre o modelo matemático desenvolvido e o modelo original	41
4.3 ABORDAGEM DE SOLUÇÃO	43
4.4 VALIDAÇÃO DO MODELO	46
5. TESTES COMPUTACIONAIS	53
5.1 DESCRIÇÃO DAS INSTÂNCIAS	53
5.2 MODELO SEM TRANSFORMAÇÃO DE PRIORIDADE (TON E DA SILVA)	54
5.3 FERRAMENTA DE <i>DESIGN</i> DE EXPERIMENTOS (<i>IRACE</i>)	55
5.4 TESTES COMPUTACIONAIS	56
6. RESULTADOS COMPUTACIONAIS	58
6.1 <i>MAKESPAN</i>	58
6.2 ANÁLISE DA FUNÇÃO OBJETIVO (FO)	59
6.3 TEMPO DE RESPOSTA POR PRIORIDADE	61

6.4 TEMPO OCIOSO (MAQUEIROS)	64
6.5 ATRASO POR PRIORIDADE.....	65
6.6 TEMPO DE ATRASO REAL E ANÁLISE DA PRIORIDADE 4	67
7. CONCLUSÃO	72
REFERÊNCIAS BIBLIOGRÁFICAS	75

1. INTRODUÇÃO

A crescente aplicação da otimização matemática está associada à constante necessidade de soluções melhores (mais rápidas e eficazes) para problemas do dia a dia da sociedade. Particularmente, em hospitais, um dos problemas existentes é o tempo de espera de pacientes para serem deslocados entre os setores do hospital.

Os hospitais são locais em que o tempo de atendimento tem grande importância. Muitas vezes, a vida dos pacientes depende da rapidez dos procedimentos de saúde em si, como da espera para serem deslocados entre consultas, cirurgias, raio-x, entre outros. Segundo Hanne et al. (2009), em alguns grandes hospitais, o serviço de transporte de pacientes é mal gerenciado e carece de coordenação do fluxo de trabalho. O trabalho de Ton e Da Silva¹ (*in press*) é um estudo inicial sobre sistemas de ressequenciamento de pacientes no ambiente hospitalar buscando minimizar tempo de espera para o deslocamento e foi utilizado como motivação para o desenvolvimento desse estudo de tempos de espera para deslocamento.

Do ponto de vista prático, sabe-se que um sequenciamento ineficaz de chamadas para transporte pode causar vários problemas dentro de um hospital. Estão inclusos: aumento do tempo de espera para o deslocamento, intervalos de tempo sem deslocamento de paciente, superdimensionamento de funcionários maqueiros para aquele período do dia possivelmente deixando funcionários ociosos. Um sequenciamento eficaz é capaz de evitar espaços ociosos de tempo e trabalho e diminuir o tempo de espera dos pacientes.

Com o propósito de evitar o sequenciamento ineficaz, este trabalho buscou a otimização da etapa de deslocamento das macas com pacientes entre unidades hospitalares (centro cirúrgico, laboratórios, UTI, enfermarias, ambulatorios, entre outros) tratando especificamente o nível de prioridade de atendimento que cada paciente recebe na chegada ao hospital. Para a resolução do sequenciamento, o problema foi tratado como o problema de máquinas trabalhando em paralelo (mesma situação do artigo que serviu de base para esse estudo).

O problema de máquinas paralelas trata da alocação de um determinado

¹ Agradecimento aos autores Ton e Da Silva pelo trabalho que serviu de base e inspiração para o estudo aqui desenvolvido.

número (n) de tarefas com tempos de execução conhecidos para serem executadas em um número (m) de máquinas funcionando em paralelo. Após a distribuição das tarefas às máquinas, o maior tempo de conclusão das tarefas, *makespan*, deve ser o mínimo possível (MULLER et al., 2002). Para Cheng et al. (1990), a grande preocupação na alocação de múltiplas máquinas é como fornecer um sequenciamento adequado de tarefas, evitando tempos ociosos, e determinar a sequência dessas tarefas em cada máquina para atingir algum objetivo prescrito. Considera-se que as máquinas sejam os conjuntos de macas e maqueiros disponíveis e as tarefas pelos pacientes, foi desenvolvido um modelo matemático para minimizar o tempo total de execução de todas as tarefas existentes (*makespan*), incluindo tempos de deslocamento e atrasos nesse deslocamento.

De acordo com Cildoz et al. (2019), a disciplina prioritária se aplica, onde, em geral, os pacientes com lesões com risco de morte são tratados antes de outros. Segundo Prado (2009), uma das estratégias de gestão, que também é usada como regra de prioridade para organizar filas de sequenciamento, é o índice de prioridade, onde o processamento é feito de acordo com o valor da prioridade atribuída ao cliente ou ao produto (paciente). Logo, cada paciente recebe um nível de prioridade na chegada e a ele é associado um tempo limite para deslocamento. Esse tempo limite será tratado como tempo para transformação entre as prioridades.

Com base nos tempos originais de espera em um hospital canadense, estes foram analisados para verificar se havia alguma possível diminuição de espera para que o paciente receba uma transformação de prioridade. Ao término desse tempo limite ocorre uma transformação para um nível mais prioritário para ser deslocado. Buscando evitar atrasos no transporte no decorrer do dia esses tempos limites foram aqui estudados e analisados. A transformação entre os níveis de prioridade tem influência direta no sequenciamento desses pacientes nas filas visto que ocorre a priorização desse paciente no modelo da alocação. Sendo assim, foi realizado um estudo sobre os tempos obtidos buscando um melhor sequenciamento dos pacientes e uma melhor programação das chamadas durante o dia resultando na diminuição do *makespan*, e por fim, no tempo total do paciente no hospital.

A principal contribuição deste trabalho, utilizando PLIM, se trata de tempos mais justos e suficientes para que, individualmente, o paciente possa esperar menos

para esse deslocamento.

1.1 OBJETIVOS

A realização deste trabalho tem o propósito de atender alguns objetivos previamente definidos. Estes objetivos são divididos em objetivo geral e objetivos específicos.

1.1.1 Objetivo Geral

Considerando as mudanças de prioridade, o objetivo do trabalho foi desenvolver um modelo matemático de PLIM para minimizar o tempo de conclusão e de atrasos do deslocamento intra-hospitalar de pacientes.

1.1.2 Objetivos Específicos

Este estudo, de forma a atender o seu objetivo geral, também pretende atingir os seguintes objetivos específicos:

- a) Expandir o modelo matemático de Ton e Da Silva (*in press*), fazendo a adaptação necessária para transformação dos níveis de prioridade;
- b) Analisar a reprogramação do sequenciamento de pacientes com base nos pacientes inseridos e/ou retirados para verificar possíveis melhorias;
- c) Baseado nos tempos originais do caso real, utilizar ferramenta de *design* de experimentos e encontrar melhor tempo para transformação de prioridades;
- d) Resolver modelo matemático com as instâncias utilizadas no trabalho citado como base. Comparar e analisar resultados encontrados entre o funcionamento atual do hospital e com as melhoras sugeridas e verificar eficiência do modelo;
- e) Realizar uma análise de sensibilidade do nível de prioridade com maior porcentagem de atrasos.

1.2 JUSTIFICATIVA DO TRABALHO

O agendamento de deslocamento de pacientes, considerando a prioridade no

atendimento, é um conceito amplamente divulgado e já usado em hospitais. A utilização de aplicativos para mensurar o tempo de espera também podem ser encontrados aqui no Brasil em aplicativos como o HNSG e SOSPS – PS. Entretanto, a modificação dos níveis de prioridades dos pacientes da fila de atendimento é uma ferramenta remotamente utilizada pelas organizações de saúde. O tempo dessa mudança de prioridade é algo que deve ser estudado e atualizado sempre que possível de acordo com o cenário em questão. Mudar a prioridade do paciente após um tempo curto de espera pode não ser a melhor alternativa, pois em pouco tempo todos os pacientes terão prioridade máxima, fato não ajudará na organização do transporte.

Um dos estudos mais próximos desenvolvidos com esse objetivo é de Aragon et al. (2018). Os autores buscam minimizar critérios de otimização da rotina hospitalar a partir de um modelo matemático, porém não consideram o sequenciamento como parte do problema desenvolvido. Outro estudo parecido foi de Majzoubi et al. (2021) verificando um problema de transporte de pacientes em tempo real do serviço de emergência em que os veículos são atribuídos aos pacientes para que possam ser transportados para os hospitais durante uma emergência. O objetivo é minimizar o tempo total de viagem de todos os veículos, satisfazendo ao mesmo tempo dois tipos de restrições de janelas de tempo.

Foi encontrado um trabalho sobre problemas intra-hospitalares com ferramentas de Pesquisa Operacional para efetuar um ressequenciamento das tarefas ao longo de um dia de trabalho dependendo de cenários e utilizando nível de prioridade, que foi o trabalho utilizado como base para esse trabalho dos autores Ton e Da Silva (*in press*). Expandindo a ideia, foi realizado o estudo com a mudança para uma prioridade maior.

Guiados pela necessidade vinda de um hospital canadense de estudar melhores tempos limite para deslocamento, esse trabalho foi desenvolvido. A busca é por aproximar cada vez mais os modelos da literatura mais próximo da realidade. Isso é feito por meio da inclusão de variáveis e parâmetros de modo que estes possam atuar no processo de ajuste do sistema tradicional de planejamento e controle da produção para o estado de um sistema dinâmico real desejado.

1.3 ESTRUTURA DO TRABALHO

O trabalho está organizado da seguinte forma:

Capítulo 2: No segundo capítulo ocorre um breve resumo dos problemas de produção e a descrição do problema de sequenciamento;

Capítulo 3: No terceiro capítulo é apresentado o levantamento bibliográfico realizado (revisão sistemática da literatura) e realiza-se uma revisão de trabalhos publicados correlatos em diferentes áreas do conhecimento;

Capítulo 4: O quarto capítulo é iniciado com o problema de transporte de pacientes intra-hospitalar, apresenta a criação do modelo matemático implementado e mostra a validação do modelo com exemplo numérico;

Capítulo 5: Neste capítulo ocorre a descrição das instâncias utilizadas, utilização de uma ferramenta de design de experimentos para auxiliar o estudo dos tempos de prioridade e como se deram os testes computacionais, quais os tipos utilizados para posterior comparação de desempenho;

Capítulo 6: Neste capítulo foram descritos os resultados encontrados após a realização dos testes, bem como a comparação do desempenho de cada grupo de testes realizado;

Capítulo 7: O último capítulo apresenta a conclusão do estudo e sugestões futuras.

1. DESCRIÇÃO DO PROBLEMA

Neste capítulo será apresentada uma breve revisão sobre problemas de produção e sequenciamento em máquinas paralelas com o objetivo de minimizar o *makespan*.

2.1 PROBLEMAS DE PRODUÇÃO

O propósito do planejamento e controle é, segundo Slack et al. (2002), “garantir que os processos da produção ocorram eficaz e eficientemente e que produzam produtos e serviços conforme requeridos pelos consumidores”. Segundo Correa et al. (2006), o sequenciamento da produção “Consiste em decidir quais atividades produtivas (ou ordens/ instruções de trabalho) detalhadas devem ser realizadas, quando (momento de início ou prioridade na fila) e com quais recursos (matérias-primas, máquinas, operadores, ferramenta, entre outros) para atender a demanda, informada, ou através das decisões do MPS”.

De acordo com Netto (2014), os problemas de sequenciamento de produção determinam uma ordem (sequência) em que as tarefas (itens) - dentre as disponíveis para o processamento - devem ser executadas em uma determinada máquina, e em que momento este sequenciamento deve ser realizado. Nesses tipos de problema de sequenciamento de produção existem alguns conceitos importantes a serem destacados para entendimento deste problema.

De acordo com a classificação de Graham et al. (1979), um problema de sequenciamento de produção é definido com base em três campos, denominados $\alpha||\beta||\gamma$; onde α denota o ambiente de produção selecionado, β descreve características referentes às tarefas ou às máquinas e γ indica a função objetivo do problema (com um único objetivo ou multiobjetivo). De acordo com Fuchigami (2013) existem alguns tipos de ambiente de produção (α):

- Máquina Única: apenas uma máquina disponível ao processamento;
- Máquinas Paralelas: várias máquinas disponíveis ao processamento, porém uma tarefa precisa passar por apenas uma das máquinas do estágio de produção;
- *Flow Shop*: existe uma sequência pré-definida de máquinas onde as tarefas necessitam passar por cada uma das m máquinas;

- *Job Shop*: cada operação possui sua própria sequência pré-definida de máquinas, não necessitam passar por cada uma das máquinas e ainda podem passar mais de uma vez pela mesma máquina;

- *Open Shop*: similar ao *Job-Shop*, entretanto as tarefas não necessitam de uma sequência própria pré-definida, podem algumas vezes seguir qualquer ordem de máquinas.

Neste trabalho o ambiente de produção (α) é de máquinas paralelas. As restrições apresentadas no campo β podem ser encontradas no capítulo 4 deste trabalho. Será visto que as tarefas têm o tempo de execução diferentes e as máquinas são idênticas. Sobre o campo γ , o modelo matemático tem como função objetivo minimizações do *makespan* (C_{max}) que é o instante de término da última tarefa a ser processada (LI et al., 2012). Este é o critério de otimização mais abordado em problemas de sequenciamento de produção, e neste trabalho também foi somada a função objetivo a minimização de atrasos dos pacientes.

2.2 MÁQUINAS PARALELAS

O problema de máquinas paralelas tem sua importância tanto do ponto de vista teórico quanto do prático. Do teórico, é uma generalização da máquina única e um caso especial de fluxo flexível. Do ponto de vista prático, é importante porque a ocorrência de recursos em paralelo é comum no mundo real, como o emprego em indústrias, agendamento de ônibus, transporte, telecomunicações, agendamento de cirurgias, de acordo com Pinedo (2016).

Para Cheng e Sin (1990), consiste no estudo de construir cronogramas de processamento para um conjunto de tarefas a fim de garantir a execução de todos os trabalhos em um período razoável. Algumas questões são tratadas: Quais trabalhos devem ser alocados para quais máquinas? Como solicitar os trabalhos de forma apropriada na sequência de processamento? A grande preocupação é como fornecer uma combinação perfeita de máquinas para trabalhos e subsequentemente determinar o processamento dos trabalhos em cada máquina para atingir algum objetivo prescrito.

Um conceito importante no estudo de máquinas paralelas é o *job*, que se trata de tarefas a serem processadas. Pode-se redefinir *job* como um pedido a ser

produzido. Outro conceito importante a ser apresentado é a preempção. A preempção, sugere que não é necessário manter uma tarefa em uma máquina até o término do processamento. É possível interromper o processamento de uma tarefa a qualquer momento, e continuar o processamento em uma máquina diferente. A quantidade de tempo de processamento que uma tarefa interrompida já recebeu não é perdida. Quando uma tarefa interrompida volta à máquina (ou a outra máquina no caso de máquinas paralelas), ela só precisa da máquina para o tempo de processamento restante. A não-preempção seria o caso de quando uma tarefa, uma vez alocada a uma máquina, permanece nela até o término de sua execução, sem interrupções, de acordo com Ji et al. (2019). Outro conceito é o de *setup* que se trata no tempo intermediário entre a execução de duas tarefas, ou tempo de preparação da máquina de acordo com Tavakkoli-Moghaddam et al. (2009).

Com base nas características, um sistema clássico de máquinas paralelas pode ser categorizado como idêntico, uniforme ou não relacionado. O cenário de máquinas paralelas idênticas é aquele em que um trabalho pode ser processado por qualquer uma das máquinas livres e o tempo de processamento é o mesmo independente da máquina alocada. Cada trabalho finalizado liberará uma máquina e deixará o sistema (CHENG e SIN, 1990).

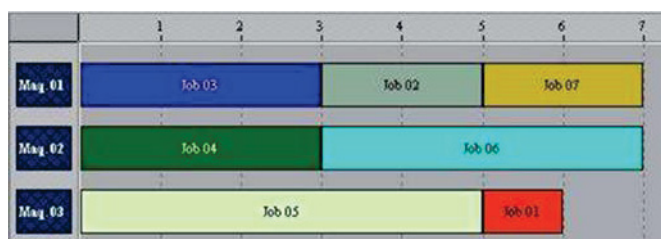
O caso de máquinas paralelas uniformes ocorre quando existe um conjunto único contendo os tempos de execução (ou finalização) das tarefas, mas os tempos de execução são alterados por um fator uniforme, dependendo da tarefa a ser atribuída a uma máquina ou à outra. O sistema de máquinas paralelas uniforme geralmente está presente quando fábricas comprem novas máquinas, mas retêm máquinas mais lentas e mais antigas, segundo Lin e Ying (2017).

O caso de máquinas paralelas não-relacionadas (*unrelated*), segundo a definição de Pinedo (2016), acontece quando n tarefas precisam ser distribuídas entre m máquinas, mas cada *job* é representado por um subconjunto de m tempos de execução, podendo ser iguais ou diferentes para cada máquina. Isso ocorre quando se fazem readequações do processo de fabricação ou quando se tem células flexíveis de manufatura, onde várias máquinas estão aptas a desempenhar diversas tarefas, porém elas podem ser mais rápidas para determinados procedimentos enquanto outras são mais rápidas para outros, não havendo possibilidade de estabelecer uma relação de velocidade.

A decisão de quais atividades produtivas (ou ordens/instruções de trabalho) devem ser realizadas, quando (momento de início ou prioridade na fila) e com quais recursos (matérias-primas, máquinas, operadores, ferramentas) para atender a demanda informada, ou por meio das decisões do plano-mestre de produção (documento que diz quais itens serão produzidos e quando cada um será produzido) é definido como *scheduling*, segundo Correa et al. (2006).

Uma ferramenta largamente utilizada para o processo de agendamento (*scheduling*) é o Gráfico de Gantt que, segundo Slack et al. (2002), “proporciona uma representação visual simples do que deveria e do que está realmente acontecendo na operação. Além disso, eles podem ser usados para pôr em teste programas alternativos”. Sua utilização é dada como artifício e ferramenta de visualização nos principais *softwares* de *scheduling*, porém não é uma ferramenta de otimização. A Figura 1 feita por Pimenta (2008) exemplifica a utilização do gráfico de Gantt em uma situação em que são processadas sete tarefas (*jobs*), em três centros de processamento (máquinas).

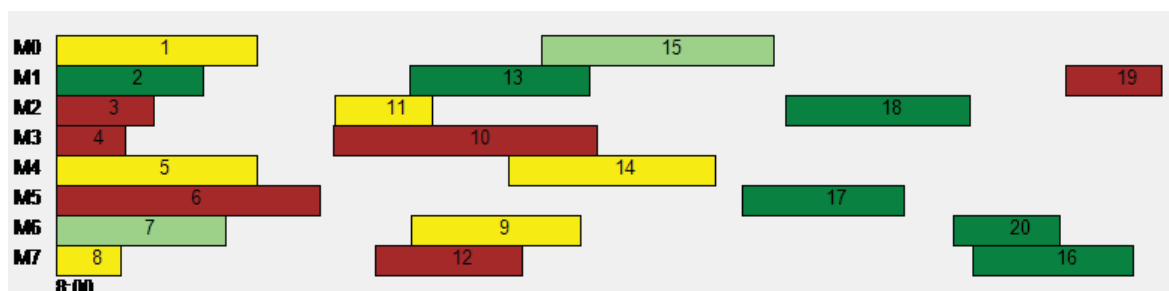
Figura 1 - Exemplo de utilização do gráfico de Gantt (Software)



Fonte: Pimenta (2008).

Nesse exemplo, observa-se a otimização das tarefas por meio do sequenciamento nos três centros de processamento, de forma clara e de fácil entendimento, onde a distância horizontal equivale ao tempo de cada *job* e o objetivo é a diminuição do tempo total de processamento, conhecido como *makespan*. Na Figura 2 é mostrado um outro exemplo do Gráfico de Gantt desenvolvido para este trabalho na linguagem de programação C#, porém com 8 máquinas e 20 *jobs*.

Figura 2 – Exemplo de utilização do gráfico de Gantt com 8 máquinas e 20 jobs



Fonte: A autora (2021).

Os problemas de sequenciamento de produção são considerados do tipo *NP-hard* (CHENG e DING, 1999), ou seja, são problemas cujo tempo computacional despendido é não determinístico polinomial. Sobre a abordagem de máquinas paralelas para resolver o problema de tempo de transformação dos níveis de prioridade no deslocamento de pacientes em macas hospitalares, foi feita uma revisão na literatura encontrada a seguir.

A respeito do deslocamento entre os diversos setores de um hospital também existe uma fila de sequenciamento para organização e controle na hora de solicitar um maqueiro para realizar o deslocamento. Para minimizar tempos de espera dos pacientes através do estudo do ressequenciamento entende-se que o tempo de deslocamento dos pacientes não pode ser modificado, bem como o instante em que o paciente entra no sistema para ser deslocado. O único fator modificável é a sequência de deslocamento dos pacientes visto que o tempo de *setup* (deslocamento em minutos com a maca vazia) entre eles é diferente, e a cada vez que o tempo limite expira esse paciente pode receber um peso diferente para ser sequenciado na frente de outros pacientes por estar aguardando a mais tempo.

2. REVISÃO DA LITERATURA

Para o estudo do tempo de transformação de prioridades, neste capítulo será apresentada uma revisão de literatura sobre o tema de máquinas paralelas e de possíveis abordagens para tratar e resolver o problema. Em seguida, outros trabalhos são descritos para a continuidade da pesquisa.

A definição inicial dos termos de busca ocorreu por meio de busca exploratória na base de dados *Web of Science*, considerando termos já conhecidos das duas áreas (Pesquisa Operacional e Hospitalar). As buscas foram feitas da seguinte maneira: filtro por tópico, somente em inglês e tipo de documento somente artigos. Ao lado das expressões de busca tem-se o número de artigos encontrados:

- I. *Patient Transportation* (132 artigos)
- II. *Queue Priority* (18 artigos)
- III. *Dynamic Queue* (61 artigos)
- IV. *Dynamic Priority Queue* (9 artigos)
- V. *Change Priority* (9 artigos)
- VI. *Hospital* (790.933 artigos)
- VII. *Model* (7.213.265 artigos)
- VIII. *Parallel Machine* (1.849 artigos)

Após essa primeira etapa, houve uma segunda busca cruzando alguns termos de busca, utilizando o operador lógico *AND*. Algumas buscas não tiveram resultados encontrados.

- I. *Hospital + Patient Transportation* (64 artigos)
- II. *Model + Dynamic Queue* (26 artigos)
- III. *Model + Patient Transportation* (30 artigos)
- IV. *Hospital + Queue Priority* (0 artigos)
- V. *Queue Priority + Patient Transportation* (0 artigos)
- VI. *Parallel Machine + Patient Transportation* (0 artigos)
- VII. *Model + Queue Priority* (9 artigos)

Foi realizada a leitura do título e resumo dos artigos encontrados em busca

única (*patient transportation, queue priority, dynamic queue, dynamic priority queue e change priority*) e dos artigos de busca com operador lógico. A partir disso, foram selecionados a quantidade de artigos a seguir para a realização da revisão sistemática do tema proposto, buscando assim identificar um *gap* desse tipo de abordagem na literatura.

A seguir são apresentados trabalhos correlatos a problemas de filas, agendamento de tarefas, tempo de espera, entre outros. Em seguida, outros trabalhos são descritos mostrando o *gap* na literatura identificado para a continuidade da pesquisa. Esses artigos foram classificados por assuntos e apresentados em ordem cronológica dentro dos seguintes temas: agendamento de tarefas, *dial-a-ride problem*, fila de prioridade, fluxo de pacientes, tempo de espera e transporte de pacientes.

3.1 AGENDAMENTO DE TAREFAS (CIRURGIAS, ATENDIMENTOS)

Sobre agendamento de tarefas foram encontrados diversos trabalhos envolvendo janelas de tempo, programação de horário para cada pacientes na realização de procedimentos hospitalares, entre outros. Joseph Adams e Egon Balas (1998) descreveram um método de aproximação para resolver o problema mínimo de produção da programação da oficina. Ele sequencia as máquinas uma a uma, sucessivamente, identificando o gargalo entre as máquinas ainda não sequenciadas. Sempre que uma nova máquina é sequenciada, todas as sequências estabelecidas anteriormente são localizadas novamente.

Spieksma et al. (1995) investigaram o problema de agendamento com distâncias de segurança que consiste em agendar trabalhos em duas máquinas paralelas sem tempo ocioso da máquina. Todo trabalho já está atribuído à sua máquina, e apenas foi preciso especificar uma ordem dos trabalhos para cada máquina. O objetivo é encontrar pedidos dos trabalhos de modo que o tempo mínimo decorrido entre dois tempos de conclusão do trabalho seja maximizado. Provaram que esse problema é geralmente NP-*hard* e forneceram algoritmos de tempo polinomial para casos especiais.

Gerchak et al. (1996) estudaram o agendamento antecipado de cirurgia eletiva quando a utilização da capacidade das salas cirúrgicas por cirurgia de emergência, bem como por procedimentos eletivos, é incerta. Novos pedidos de reservas de cirurgia eletiva chegam todos os dias. O problema enfrentado pelo hospital no início

de cada dia é quantas solicitações adicionais de cirurgia eletiva devem ser atribuídas para esse dia. Sendo assim, foi fornecido um modelo de programação dinâmica estocástica para esse problema agregado de agendamento antecipado.

Sciomachen et al. (2005) sugeriram modelos de simulação de eventos discretos e usados para avaliar índices de desempenho de interesse relacionados ao agendamento de atividades da sala de operações para departamentos cirúrgicos; em particular, a atenção está concentrada na produtividade das enfermarias em termos de taxa de utilização, produtividade e excedentes. As sequências das atividades cirúrgicas são analisadas de acordo com diferentes regras de prioridade, derivadas da teoria clássica de programação.

Lamiri et al. (2008) criaram um modelo estocástico para o planejamento de salas cirúrgicas com dois tipos de demanda por cirurgia: cirurgia eletiva e cirurgia de emergência. O problema de planejamento consiste em atribuir casos eletivos a diferentes períodos em um horizonte de planejamento, a fim de minimizar a soma dos custos eletivos relacionados ao paciente e dos custos de horas extras das salas de cirurgia. Em seguida, proposto um método de otimização de Monte Carlo combinando simulação de Monte Carlo e Programação Inteira Mista.

Majzoubi et al. (2021) estudaram um problema de transporte de pacientes em tempo real do serviço de emergência em que os veículos são atribuídos aos pacientes para que possam ser transportados para os hospitais durante uma emergência. O objetivo é minimizar o tempo total de viagem de todos os veículos, satisfazendo ao mesmo tempo dois tipos de restrições de janela de tempo. A primeira exige que cada veículo do EMS chegue ao local de um paciente dentro de um período especificado. O segundo exige que o veículo chegue ao hospital designado dentro de outra janela de tempo.

O próximo assunto será o problema de projetar rotas e horários de veículos para cada situação dentro do hospital.

3.2 DIAL-A-RIDE PROBLEM

O problema DARP consiste em projetar rotas e horários de veículos para n usuários que especificam solicitações de coleta e entrega entre origens e destinos. Os autores Cordeau e Laporte (2003) descreveram uma heurística de pesquisa tabu para o problema de discagem automática com as seguintes características: os

usuários especificam solicitações de transporte entre origens e destinos. Eles podem fornecer uma janela de tempo na hora de partida ou chegada desejada. O transporte é fornecido por uma frota de veículos com base em depósito comum. O objetivo é projetar um conjunto de rotas de veículos de menor custo, capazes de acomodar todos os pedidos. As restrições laterais estão relacionadas à capacidade do veículo, duração da rota e tempo máximo de viagem de qualquer usuário.

No problema de discagem, os usuários formulam solicitações de transporte de uma origem específica para um destino específico. O transporte é realizado por veículos que fornecem um serviço compartilhado. O problema consiste em projetar um conjunto de rotas de veículo com custo mínimo que satisfaçam as restrições de capacidade, duração, janela de tempo, emparelhamento, precedência e tempo de viagem. Este trabalho introduz um problema de programação linear inteira mista e um algoritmo de ramificação e corte, desenvolvido por Cordeau (2006).

O objetivo neste trabalho foi planejar um conjunto de m rotas de veículos de custo mínimo capazes e acomodá-los da melhor maneira possível, sob um conjunto de restrições. O exemplo mais comum surge em transporte porta a porta para idosos ou pessoas com deficiência. O objetivo do artigo de Cordeau e Laporte (2007) foi de revisar a literatura científica sobre o DARP, destacando as principais características do problema e um resumo dos modelos e algoritmos mais importantes.

Molenbruch et al. (2017) fizeram a generalização de um problema biobjetivo de discagem, incorporando características da vida real do transporte de pacientes. Um algoritmo de busca local multidirecional foi desenvolvido para resolver esse problema. Um *local search* foi integrado a uma estrutura de descida de vizinhança variável que aplica um princípio inteligente da lista de candidatos para reduzir o tempo de computação. Além disso, foi proposto um novo procedimento de agendamento, construindo agendamentos que minimizam o tempo total de viagem do usuário.

Diversos trabalhos tiveram o mesmo tema geral envolvendo filas de prioridades e a seguir serão expostos alguns trabalhos.

3.3 FILA DE PRIORIDADE

As filas envolvendo prioridade no atendimento são largamente utilizadas atualmente, bem como na fila de atendimento em hospitais. Kofman e Lippman (1981) fizeram um artigo considerando uma fila $M / M / 1$ com dois tipos de clientes: clientes

prioritários e clientes regulares. Eles chegam à instalação de serviço de acordo com dois fluxos independentes de Poisson e formam uma fila única de acordo com a ordem em que chegam. O servidor seleciona os clientes a serem atendidos para minimizar o custo médio esperado por unidade de tempo de operação do sistema. Mostraram que, sempre que o número de clientes regulares ignorados em uma promoção, vezes os custos esperados de retenção por cliente prioritário por período de serviço, é maior ou igual a R , a promoção é estritamente ideal. Além disso, para cada estado existe um valor de R – com R excedendo o número de clientes regulares ignorados em uma promoção, multiplicado pelos custos esperados de manutenção por cliente prioritário por período de serviço – para o qual a promoção é ideal.

Siddharthan et al. (2010) investigaram o aumento do custo do tempo de espera imposto à sociedade devido ao uso inadequado do departamento de emergência pelos pacientes, buscando atendimento não emergencial ou primário. Propõe um modelo econômico simples para ilustrar o efeito desse uso indevido em um hospital público ou sem fins lucrativos. Este trabalho fornece evidências de que pacientes não emergenciais contribuem para longos atrasos no pronto-socorro para todas as classes de pacientes e propõe um modelo de enfileiramento prioritário para reduzir o tempo médio de espera.

Mattson (1999) estudaram um sistema de processamento de dados com várias filas para priorizar solicitações de E / S para um dispositivo de armazenamento, perceberam que a prioridade das filas para manutenção é dinamicamente ajustável, em oposição a cada fila com uma prioridade fixa. O ajuste dinâmico da prioridade de uma fila permite a prioridade da *queue* com solicitações “travadas” a serem levantadas para que as “travadas” solicitações podem ser atendidas muito rapidamente.

Agrawal (2010) apresentou uma abordagem teórica de filas para uma classe geral de sistemas de saúde. Foram derivadas novas expressões para avaliar o efeito de interrupções no serviço e usado o modelo resultante para aproximar o tempo de fluxo do paciente. Estudou-se o impacto prejudicial das interrupções de serviço nos tempos de fluxo do paciente. Além disso, um modelo de otimização foi apresentado para determinar o número ideal de médicos/prestadores de serviços necessários durante uma sessão de serviço baseado em número de pessoal, taxas de chegada de pacientes e horários de atendimento.

Pajouh e Kamath (2010) concluíram que, em um sistema hospitalar, aumentar

a utilização de recursos para reduzir custos e diminuir o tempo de espera dos pacientes para fornecer atendimento oportuno e melhorar a satisfação dos pacientes são objetivos importantes, mas conflitantes. Os modelos de enfileiramento podem fornecer avaliações razoavelmente precisas do desempenho do sistema. Os autores revisaram e categorizaram a literatura na tentativa de motivar novas pesquisas na aplicação de modelos de filas no domínio da saúde.

Hagen et al. (2013) examinaram vários modelos de filas diferentes para UTI e os efeitos nos tempos de espera, utilização, taxas de retorno, mortalidade e número de pacientes atendidos. Um modelo de simulação baseado em sistema é construído para capturar todos os casos possíveis de fluxo de pacientes após a admissão na UTI.

Ferrand et al. (2018) investigaram duas abordagens distintas para enfrentar esses desafios de capacidade em departamentos de emergência: via rápida e fila de prioridade dinâmica. Usando uma simulação empírica, compararam o impacto de diferentes abordagens de alocação e priorização de recursos no tempo de permanência do paciente, incluindo o sistema existente no departamento de emergência, via rápida com dedicação estrita e parcial e a possibilidade de tempos de serviço mais curtos e menos variáveis, e versões da fila de prioridade dinâmica proposto usando priorização dinâmica simples.

3.4 FLUXO DE PACIENTES

Foi projetado um sistema de planejamento para o transporte intra-hospitalar baseado em computador, o Opti-TRANS ©, que suporta todas as fases do fluxo de transporte, incluindo reservas de viagens, envio de solicitações de transporte e monitoramento e relatórios de viagens em tempo real. A metodologia que Hanne et al. desenvolveram em 2009 para resolver o problema de otimização subjacente – um problema dinâmico de discagem com restrições específicas de hospitais, baseia-se em métodos heurísticos rápidos para garantir o fornecimento eficiente e oportuno de transportes. Foi ilustrado o forte impacto do Opti-TRANS no desempenho diário do serviço de transporte de pacientes de um grande hospital alemão.

Jacobson et al. (2013), após estudos, perceberam que as organizações de assistência médica enfrentam pressões cada vez maiores para oferecer atendimento de qualidade, ao mesmo tempo em que enfrentam custos crescentes, menores reembolsos e novas demandas regulatórias. O artigo fornece uma visão geral dos

aplicativos de modelagem de simulação de eventos discretos para clínicas de saúde e sistemas integrados de saúde (por exemplo, hospitais, ambulatórios, departamentos de emergência e farmácias) nos últimos quarenta anos.

Bhattacharjee e Ray (2014) fizeram uma análise dos processos hospitalares essenciais para o desenvolvimento de métodos, políticas e ferramentas de decisão aprimorados para melhorar o desempenho geral do sistema hospitalar. Também foi fornecida uma estrutura genérica para modelagem de fluxo de pacientes e análise de desempenho de sistemas hospitalares que pode servir como um guia para os profissionais que lidam com tipos semelhantes de problemas para melhorar a prestação de cuidados de saúde.

3.5 TEMPO DE ESPERA

Samaha et al. (2003) desenvolveram um artigo com um modelo de simulação das operações no DE do Sistema de Saúde Cooper. Devido à grande quantidade de variabilidade que pode ocorrer em um departamento de emergência, o *Cooper Health System* buscou o uso de simulação para ajudar a avaliar suas operações e possíveis soluções para seus problemas. O objetivo do trabalho era que a *Rockwell Automation* (Arena) criasse um modelo que descrevesse as operações atuais e avaliasse possíveis alternativas para reduzir o tempo de permanência. O modelo produziu resultados que economizaram tempo considerável ao hospital e ajudaram a evitar custos elevados.

Van Essen et al. (2012) verificaram uma dificuldade hospitalar que se trata do agendamento de cirurgias de emergência o mais rápido possível, porque adiá-las geralmente aumenta o risco de complicações de um paciente. Foram discutidos vários métodos de solução exata e heurística para esse novo tipo de problema agendando os “momentos de interrupção” da maneira mais uniforme possível. Para estimar esse efeito e investigar a robustez dos cronogramas criados, realizou-se um estudo de simulação.

Tan et al. (2013) diante de um problema de aglomeração em departamento de emergência, forneceram uma estrutura integrada para gerenciar filas dinamicamente a partir de perspectivas de demanda e fornecimento. Apresentou-se estratégias inteligentes de priorização dinâmica de pacientes para gerenciar a demanda simultaneamente com políticas dinâmicas de ajuste de recursos para

gerenciar o fornecimento. Verificou-se por meio de uma simulação que essa estrutura melhora o tempo de permanência dos pacientes no pronto-socorro sem restringir a demanda.

3.6 TRANSPORTE DE PACIENTES

Dershin (1993) descreveu um método para organizar e equipar um sistema de transporte de pacientes em todo o hospital de maneira a minimizar o tempo de espera do paciente. O método deriva de um projeto de melhoria da qualidade e inclui um sistema de comunicação centralizado, um modelo de filas para determinar os requisitos de pessoal e um sistema computadorizado de coleta e monitoramento de dados. Um estudo de caso demonstra melhorias nos tempos de espera, custos e satisfação do cliente.

Segev et al. (2012) desenvolveram uma ferramenta de simulação para permitir que os tomadores de decisão abordem simultaneamente questões de planejamento e avaliação sobre o transporte de pacientes. Codificaram uma ferramenta de simulação estocástica para um modelo generalizado, tratando todos os fatores que contribuem para o processo. Utilizaram o modelo para análises de sensibilidade do número de elevadores dedicados, transportadores dedicados, velocidade do transportador e a hora de início do processo planejado no atraso das partidas de salas de cirurgia e o número de casos com atrasos graves.

Jin et al. (2015) desenvolveram um modelo de programação inteira mista para selecionar a localização e o fluxo da rede de uma clínica no local para apoiar o tratamento de primeiros socorros próximo a áreas de desastre, bem como um modelo de hospital geral como uma função objetivo para maximizar o número de pacientes cuja probabilidade de sobrevivência excede o nível marginal.

Para Fröhlich von Elmbach et al. (2019), o transporte de pacientes dentro de um hospital envolve três grupos principais: os gerentes do hospital, que devem garantir um sistema de transporte eficiente, com carregadores suficientes, para que atrasos de cirurgias ou exames sejam (principalmente) evitados; os pacientes, para que os tempos de espera antes e depois das consultas não devem ser muito longos; e os carregadores, que realizam o transporte físico dos pacientes dentro do hospital. O artigo integra as necessidades dos três grupos mencionados em um modelo matemático, resolvido com a ajuda de um algoritmo de pesquisa tabu personalizado.

No estudo de Bonte et al. (2019) é visto que a pressão financeira contínua sobre os hospitais os obriga a repensar vários fluxos de trabalho. Tendo como foco otimizar os custos de transporte hospitalar, foi discutida uma plataforma de autoaprendizagem que aprende as causas dos atrasos no transporte, a fim de evitar esses tipos de atrasos no futuro. Uma vez aprovado, o sistema calculará quanto mais tempo deve ser atribuído a esses transportes para evitar atrasos futuros. Como resultado, o agendamento do transporte do paciente pode ser otimizado automaticamente, enquanto o gerenciamento permanece sob controle total do processo.

Beaudry et al. (2010) tiveram o objetivo de fornecer um serviço de transporte eficiente e atempado aos pacientes entre vários locais de um campus hospitalar. Os pedidos de transporte chegam de uma forma dinâmica e a metodologia de solução deve, portanto, ser capaz de inserir rapidamente novos pedidos nas atuais rotas de veículos. É proposto um procedimento heurístico de duas fases capaz de lidar com as suas muitas características (inserção simples para gerar uma solução viável, que é melhorada na segunda fase com um algoritmo de pesquisa tabu). Em particular, conseguiu reduzir os tempos de espera dos pacientes, utilizando menos veículos.

Após a descrição desses artigos, pode-se entender que não foram encontrados problemas intra-hospitalares com ferramentas de Pesquisa Operacional para efetuar um ressequenciamento das tarefas ao longo de um dia de trabalho, sabendo que o trabalho mais próximo é de Aragon et al. (2018). Os autores buscam minimizar critérios de otimização da rotina hospitalar a partir de um modelo matemático, porém não consideram o sequenciamento como parte do problema desenvolvido.

O trabalho mais atual foi o de Majzoubi et al. (2021) em que estudaram um problema de transporte de pacientes em tempo real do serviço de emergência em que os veículos são atribuídos aos pacientes para que possam ser transportados para os hospitais durante uma emergência. O objetivo é minimizar o tempo total de viagem de todos os veículos, satisfazendo ao mesmo tempo dois tipos de restrições de janela de tempo. A primeira exige que cada veículo do EMS chegue ao local de um paciente dentro de um período específico. O segundo exige que o veículo chegue ao hospital designado dentro de outra janela de tempo.

O trabalho aqui desenvolvido se trata de uma junção dos temas encontrados:

agendamento de tarefas, DRP, fila de prioridade, fluxo de pacientes, tempo de espera e transporte de pacientes. A apresentação dos artigos apresenta os diversos artigos no tema de transporte de pacientes em filas, porém não foram encontrados trabalhos utilizando a transformação do nível de prioridades. Outro aspecto a ressaltar é o uso de heurísticas para melhoria no tempo computacional é uma realidade que auxilia muito na resolução de problemas de pesquisa operacional, bem como o aumento do uso de parâmetros e variáveis deixando o problemas mais próximo da realidade e buscando o bem-estar do paciente.

4. O PROBLEMA DE TRANSPORTE DE PACIENTES DENTRO DE HOSPITAIS

Neste capítulo é apresentado o problema de transporte de pacientes dentro de hospitais, bem como está descrita a formulação matemática do modelo desenvolvido respeitando a restrição adicional de transformação de prioridades, bem como o pseudocódigo em fluxograma. Baseado em um modelo matemático já existente no trabalho de Ton e Da Silva (*in press*), foi realizada a expansão do modelo matemático para a transformação de nível de prioridade. Por fim, nesse capítulo, o modelo foi validado com um exemplo numérico.

4.1 DEFINIÇÃO DO PROBLEMA

O problema desse estudo consiste no tempo de espera para o transporte de pacientes. Após o tempo limite ocorre uma mudança na prioridade do paciente para iniciar o deslocamento. A transformação de prioridade sempre irá ocorrer de uma prioridade menos urgente para uma prioridade mais urgente.

A respeito dos níveis de prioridade no atendimento e consequente deslocamento, existe uma escala canadense de triagem ou CTAS. Se trata de uma ferramenta de triagem e foi introduzida em 1999 em departamentos de emergência do Canadá. Permite a classificação do paciente com base na urgência do seu estado de saúde (risco e intensidade dos sintomas). A escala é composta por 5 níveis de acuidade que correspondem a uma cor e tempo-alvo, assim descritos:

- Nível 1: reanimação – azul: imediato;
- Nível 2: emergente – vermelho: 15 minutos;
- Nível 3: urgente – amarelo: 30 minutos;
- Nível 4: menos urgente – verde: 60 minutos;
- Nível 5: não urgente – branco: 120 minutos;

O estudo abrange um hospital canadense o qual utiliza 4 níveis. Será visto a seguir que foram utilizados 3 cenários para testes podendo ser emergência, consultórios, UTI's, de acordo com Ton e Da Silva (*in press*). Esses cenários também podem ser entendidos como 3 horários diferentes do dia, o que causa uma diferença na quantidade de pacientes em cada prioridade por conta do nível de serviço.

Na chegada do paciente ao sistema cada paciente recebe um nível de prioridade no atendimento. Cada nível de prioridade tem um tempo máximo pré-estabelecido em que o paciente deve ser deslocado, que se neste trabalho será tratado como tempo de transformação de prioridades. Após esse tempo de transformação de prioridades começa a ser contado atraso no deslocamento e no próximo sequenciamento esse paciente terá uma prioridade mais urgente para ser alocado na fila de deslocamento. A função objetivo desse problema contabiliza esses atrasos para todos os níveis de prioridade, bem como busca minimizar o tempo de entrega do último paciente do dia.

Baseado nos tempos utilizados atualmente no hospital, foram estudados os tempos ideais de transformação para cada nível de prioridade, fazendo com que o paciente receba uma prioridade maior ao término de um tempo limite. Esse tempo limite foi testado e examinado para evitar atrasos no transporte no decorrer do dia, minimizando assim o *makespan*.

Figura 3 - Relação entre as características de um problema de produção e um problema intrahospitalar

PROBLEMA DE PRODUÇÃO	PROBLEMA INTRA-HOSPITALAR
Tarefas	Atividades
Máquinas	Circulantes
Tempo de <i>setup</i> entre as tarefas	Tempo de acesso do circulante às atividades
Tempo de processamento da tarefa	Tempo de execução da atividade
Data de lançamento da tarefa ao sistema	Data de liberação para o atendimento ao paciente
Prazo de entrega da tarefa	Prazo para a finalização do atendimento

Fonte: Da Silva (2020).

A Figura 3 anteriormente citada apresenta o quadro elaborado por Da Silva (2020), que faz um comparativo entre o ambiente de produção e o ambiente hospitalar. Apresenta os termos e características equivalentes nos dois problemas que deram base para o desenvolvimento desse trabalho. A atividades (*jobs*) são os pacientes que serão deslocados por cada maqueiro/circulante (máquina). O tempo de *setup* entre tarefas seria o tempo de deslocamento com a maca vazia entre os pacientes. A data de lançamento da tarefa ao sistema como o instante de liberação do paciente no sistema para ser deslocado, bem como o prazo para entrega da tarefa sendo o prazo para entregar o paciente em seu local de destino.

Em resumo, busca-se a redução do tempo de espera do paciente (*job*) no

transporte intra-hospitalar realizado por macas (máquinas). Foi desenvolvido um modelo matemático com o objetivo de encontrar o tempo de transformação de prioridade de acordo com cada cenário de atendimento do hospital. Essa atualização periódica é fundamental para a tomada de decisão do atendimento, dando oportunidade para um deslocamento mais rápido e eficaz.

4.2 MODELO MATEMÁTICO DE PROGRAMAÇÃO LINEAR INTEIRA MISTA

São necessários diversos parâmetros e variáveis para ter um modelo mais diversificado e com a proximidade da realidade esperada, por exemplo, incluindo-se variáveis de atraso no modelo, utilizando prioridades para atendimento, entre outros. Levando-se em consideração o cenário de uma modelagem para deslocamento hospitalar, é necessário que o mesmo considere o tempo de *setup* entre os transportes, a demanda de deslocamento de pacientes no hospital em função do horário do dia, o número de macas disponíveis, as distâncias a serem percorridas de deslocamentos.

Neste estudo, para a construção do modelo, foram utilizados dados reais de um hospital canadense. Esses dados sofreram um tratamento prévio e retirada das informações não utilizadas, sendo mantidos os horários dos deslocamentos, tempos de processamento, locais de origem e destinos para cada paciente, nível de prioridade na entrada do sistema. O problema de produção abordado é desenvolvido por um modelo de PLIM e tem como objetivo minimizar o *makespan* e o atraso total considerando a transformação de prioridades em várias máquinas. Antes da descrição do modelo, as seguintes hipóteses para o problema são pressupostas:

- Existem n pacientes (*jobs*) a serem transportados;
- Existem m macas (máquinas) para realizar os transportes
- As m macas contém m maqueiros e esses são considerados homogêneos (quando é dito maca no presente trabalho, isso refere-se ao conjunto maca maqueiro);
- Cada paciente tem um horário de liberação para ser transportado (horário em que o chamado foi adicionado ao sistema);
- Não há preempção (*preemption*) durante o sequenciamento, ou seja, as tarefas não podem ser interrompidas durante sua execução;

- Considera-se uma tarefa fictícia zero (0) que precede imediatamente a primeira tarefa de cada maqueiro (máquina).
- No trabalho dos autores Ton e Da Silva (*in press*) foram testadas 4 políticas para o ressequenciamento, isto é, qual o momento certo para fazê-lo. A política 1 dizia que a reprogramação é desencadeada sempre que chega um novo pedido. Essa foi a política que teve melhor desempenho realizando o ressequenciamento a cada entrada de novo paciente no sistema, respeitando também o número máximo de maqueiros (com 8 pacientes haveria um ressequenciamento novamente). A necessidade de ressequenciar com frequência existe pelo fato de ser um hospital, então não se sabe os chamados futuros. Cada vez que um paciente entra no sistema faz-se necessário incluí-lo naquele instante ao sequenciamento.

4.2.1 Formulação do Modelo Matemático

O modelo contém os seguintes conjuntos:

- I : conjunto de pacientes $I = \{0, \dots, n\}$;
- J : subconjunto de I , $J = I \setminus \{0\}$;
- K : conjunto de transportadores, $K = \{1, \dots, m\}$;
- W : conjunto de prioridades, $W = \{1, 2, 3, 4\}$.

As variáveis do modelo contêm os seguintes índices:

- i, j, h : índices para os pacientes, sendo $i, h \in I$ e $j \in J$;
- k : índice para o maqueiro, sendo $k \in K$;
- w : índice para os níveis de prioridades no deslocamento intra-hospitalar de acordo com a urgência de cada paciente do menos ao mais urgente, sendo $w \in W$.

Os parâmetros de entrada, inteiros e não negativos, analisados nesta dissertação são apresentados a seguir:

- p_j : tempo de deslocamento (processamento) do paciente j em minutos, onde $j \in J$.
- r_j : instante de liberação do deslocamento do paciente j em minutos;

- s_{ij} : tempo de *setup* (tempo de deslocamento entre o local de destino do paciente i para o local de origem do paciente j sendo i o antecessor imediato a j). Deslocamento com a maca vazia em minutos considerando a velocidade de todos os maqueiros igual;
- T_{wj} : tempo limite para iniciar o deslocamento do paciente j no nível de prioridade w . Após esse limite, o paciente sofre uma transformação de prioridade do nível w aumentando-se sua urgência e priorizando esse paciente no sequenciamento para deslocamento.
- M : número suficientemente grande.

As variáveis de decisão são:

- C_{jk} : instante de término do deslocamento do paciente j pelo maqueiro k ;
- C_{max} : tempo máximo de conclusão dos deslocamentos ou *makespan*, onde $C_{max} = \max_{(j,k)} C_{jk}$;
- $x_{ijk} = 1$, se o paciente i é imediatamente antecessor ao paciente j na maca k ; $x_{ijk} = 0$, caso contrário.
- ε_{jw} : tendo ciência do tempo limite para transformação de prioridade, se trata do tempo excedente (atraso), em minutos, até começar o deslocamento do paciente j com nível de prioridade w que o paciente chega ao sistema;

Tendo como base o exposto, foi desenvolvido o modelo exato de programação linear inteira mista levando em consideração a transformação de prioridades buscando diminuir atrasos e possíveis tempos ociosos de trabalho. A função objetivo e as restrições do problema de sequenciamento aplicado ao transporte intra-hospitalar com a mudança de prioridade são apresentadas no seguinte modelo de programação linear mista:

$$\min Z = C_{max} + \sum_{j=1}^n \sum_{w=1}^w \varepsilon_{jw} \quad (1)$$

Sujeito a:

$$\sum_{k=1}^m \sum_{\substack{i=0 \\ i \neq j}}^n x_{ijk} = 1, \quad \forall j \in J \quad (2)$$

$$\sum_{k=1}^m \sum_{\substack{j=1 \\ j \neq i}}^n x_{ijk} \leq 1, \quad \forall i \in I \setminus \{0\} \quad (3)$$

$$\sum_{j=1}^n x_{0jk} \leq 1, \quad \forall k \in K \quad (4)$$

$$\sum_{\substack{h=0 \\ h \neq i, j}}^n x_{hik} \geq x_{ijk}, \quad \forall i, j \in J, \quad i \neq j, \quad \forall k \in K \quad (5)$$

$$C_{jk} - C_{ik} \geq s_{ij} + p_j + (x_{ijk} - 1) * M \\ \forall j \in J, \forall i \in I \setminus \{0\}, i \neq j, \forall k \in K \quad (6)$$

$$C_{max} \geq C_{jk}, \quad \forall j \in J, \forall k \in K \quad (7)$$

$$C_{jk} - p_j - r_j \leq T_{jw} + \varepsilon_{jw}, \\ \forall j \in J, \forall k \in K, \forall w \in W \quad (8)$$

$$C_{jk} \geq (r_j + p_j + s_{ij}) * x_{ijk}, \\ \forall i \in I, \forall j \in J, \forall k \in K \quad (9)$$

$$C_{jk} \geq 0, \forall j \in J, \forall k \in K \quad (10)$$

$$C_{0k} = 0, \quad \forall k \in K \quad (11)$$

$$x_{ijk} \in \{0,1\}, \forall i \in I, \forall j \in J, i \neq j, \forall k \in K \quad (12)$$

A função objetivo é responsável pela minimização do *makespan* com coeficiente 1 e o somatório dos atrasos por paciente e por prioridade tendo cada prioridade um peso relacionado de coeficiente (1). O conjunto de restrições (2) garante que cada atividade, associada a um makeiro, possui uma única atividade imediatamente antecessora. O conjunto de restrições (3) afirma que existe, no máximo, uma atividade imediatamente sucessora a outra, caso ambas sejam realizadas pelo mesmo makeiro. O conjunto de restrições indicada em (4) informa

que a atividade fictícia, responsável pela inicialização do sistema, pode ser sucedida por no máximo uma atividade para cada maqueiro. O conjunto de restrições (5) assegura que as atividades predecessoras e sucessoras entre si são efetuadas pelo mesmo maqueiro.

Certifica-se no conjunto de restrição (6) que, considerando dois pacientes imediatamente seguidos e transportados pelo mesmo maqueiro, o instante de término do deslocamento do paciente j deve ocorrer após a soma entre: o instante de término de deslocamento do paciente i , o tempo de *setup* entre eles, e o tempo de deslocamento do paciente j . O conjunto de restrição (7) diz que o tempo máximo de conclusão das atividades deve ser maior ou igual ao instante de término de cada atividade, para todos os maqueiros. O conjunto de restrição (8) afirma que o instante de término de deslocamento do paciente j pelo maqueiro k subtraído do instante de liberação desse mesmo paciente (momento em que a chamada entrou no sistema) e do tempo de deslocamento do paciente seja menor ou igual ao tempo de transformação de prioridade somado a uma penalidade caso isso não seja respeitado.

O conjunto de restrição (9) determina que o instante de término da atividade seja maior que a soma do instante de liberação, tempo de deslocamento e o *setup* entre os pacientes imediatamente seguidos. Em (10) é estabelecido que o instante de término da atividade determina-se que o instante de término das atividades é não-negativo. O conjunto (11) diz que o instante de término do paciente fictício deve ser 0 e o conjunto de restrição (12) indica o domínio da variável de decisão x_{ijk} .

4.2.2 Diferenças entre o modelo matemático desenvolvido e o modelo original

A grande diferença entre os dois modelos é a adição do ressequenciamento ao fim do tempo de transformação de prioridade, realizando assim uma transformação na prioridade original do paciente. O parâmetro T_{jw} considera o tempo de cada nível de prioridade por paciente, fator esse que faz diferença e torna o problema mais próximo da realidade. Como falado anteriormente, trabalhos considerando uma atualização de prioridades juntamente com ressequenciamento não existem na literatura. Serão apresentadas nesse momento as diferenças e restrições que foram sofreram modificações:

A função objetivo original tem como coeficiente $\lambda \in (0,1)$ e representa peso λ

e $\lambda-1$, enquanto no modelo atual o *makespan* tem coeficiente 1 e os atrasos em cada nível de prioridade recebem um coeficiente diferente de acordo com cada nível. O conjunto de restrição 11 original diz que *lateness* é calculado pela diferença entre o instante de término da atividade j e seu prazo de finalização estipulado (*due date*).

$$T_j \geq C_{jk} - d_j, \quad \forall j = 1, \dots, n, \quad \forall k = 1, \dots, m \quad (11)$$

$$C_{jk} - p_j - r_j \leq T_{jw} + \varepsilon_{jw}, \\ \forall j = 1, \dots, n, \forall k = 1, \dots, m, \quad \forall w = 1, 2, 3, 4 \quad (8)$$

No modelo desenvolvido e apresentado anteriormente, o conjunto de restrição 8 tem a variável ε_{jw} incluída, bem como a adição de um índice no *lateness*: o nível de prioridade (w). Sendo assim, o conjunto de restrição 8 busca garantir que o instante de término de deslocamento do paciente subtraído do tempo de deslocamento de j e o instante de liberação do paciente devem ser menores ou iguais ao tempo de transformação de prioridade, que é um parâmetro de entrada. Desta forma o que ultrapassar o tempo de transformação será adicionado como atraso desse paciente nessa prioridade.

No trabalho de Ton e Da Silva (*in press*) citado anteriormente foram testadas algumas políticas de ressequenciamento dos pacientes. As políticas são diversos modos de programar em qual momento haverá um novo sequenciamento. Na política 1 o ressequenciamento é desencadeado sempre que chega um novo pedido. Na política 2 os pedidos ficam em uma fila de espera e a reprogramação é acionada com a chegada de um pedido urgente. A política 3 é semelhante à 2, mas o ressequenciamento é desencadeado periodicamente a cada tempo determinado em vez de em momentos variáveis, ou à chegada de um pedido urgente. Por fim na política 4 é desencadeada uma reprogramação sempre que um pedido é completado e considera apenas encomendas mais antigas na fila.

Foram analisados esses 4 tipos de políticas e a política 1 teve o melhor desempenho. Essa política 1, que realiza um ressequenciamento a cada novo paciente no sistema, será utilizada para o estudo e serão analisados os cenários do hospital e os tempos de espera para cada nível de prioridade.

4.3 ABORDAGEM DE SOLUÇÃO

A Figura 4 mostra como funciona o modelo e o ressequenciamento dos pacientes por meio de um fluxograma do código.

Figura 4 - Fluxograma de funcionamento do modelo matemático desenvolvido



Fonte: A autora (2021).

A primeira fase se trata da leitura dos dados na instância em txt. Após essa leitura são criados dois objetos de estrutura idêntica: *SolucaoAtual* e o *SolucaoFinal* que são vetores de listas. A quantidade de linhas é o número de maqueiros k e as colunas têm tamanho indeterminado recebendo a quantidade de *jobs* encaminhada àquele maqueiro. A estrutura *SolucaoAtual* é atualizada a cada ressequenciamento e o *SolucaoFinal* vai sempre recebendo e adicionando *jobs* até o fim apresentar o sequenciamento final. Foi criado no programa um contador para representar o passar do tempo. Logo, com o andamento do “relógio” a cada segundo é feita uma verificação

dos instantes de liberação de cada *job*. Caso o instante de liberação seja menor ou igual ao instante do relógio, esse paciente entra para o vetor *JobsParaConsiderar* (lista de *jobs* liberados para serem considerados no sequenciamento).

Para a primeira rodada, o *job* 0 (fictício) ocupa a primeira posição em todos os maqueiros, bem como na *SolucaoAtual*. Esse mesmo *job* também ocupa a primeira posição na estrutura *SolucaoFinal* dando início ao sequenciamento. Esse *job* fictício tem instante de liberação, tempo de deslocamento e prioridade iguais a 0. Ele serve como apoio para a construção do modelo matemático.

A entrada e/ou saída de *jobs* da lista *JobsParaConsiderar* representa a fase “Atualiza lista de Jobs” do fluxograma. Quando obedecida a verificação, ou seja, existem *jobs* com instante de liberação igual a ao instante do relógio, incluem-se novos *jobs* nessa lista. A próxima pergunta a ser respondida é: “Existem novos *jobs* nessa lista *JobsParaConsiderar*?”. Se a resposta for sim, algumas etapas são seguidas: primeiramente são retirados todos os *jobs* da estrutura *SolucaoAtual*, o modelo matemático é montado e executado gerando um sequenciamento de pacientes como resultado. Com esse resultado é montada a estrutura *SolucaoAtual* com todos os *JobsParaConsiderar* visto que todos estão aptos para iniciar o deslocamento.

Verifica-se se o deslocamento de cada paciente já iniciou. Caso já tenha iniciado esse *job* é inserido na estrutura *SolucaoFinal* (como o caso de máquinas paralelas aqui estudado considera-se sem preempção, após o maqueiro começar a se deslocar para buscar um paciente ele não pode sofrer alteração de percurso, deve realizar o deslocamento completo). Após inserido na *SolucaoFinal* ele não pode mais ser modificado. Caso a resposta para a pergunta “Existem novos *jobs* nessa lista *JobsParaConsiderar*?” for negativa simplesmente ocorre a atualização da *SolucaoFinal* visto já existe um sequenciamento anterior que será obedecido caso não haja entrada de novo paciente. Verifica-se o instante do relógio e caso algum paciente anteriormente programado para um maqueiro tenha seu deslocamento iniciado, ocorre a atualização desse paciente para a *SolucaoFinal*.

Acada instante (atualização do relógio) os *jobs* vão sendo inseridos ao sistema, a lista *JobsParaConsiderar* vai sendo atualizada recebendo e perdendo *jobs*, bem como a estrutura de *SolucaoAtual*. De acordo com a política 1 que obteve melhor resultado no artigo de Ton e Da Silva (*in press*), os pacientes vão sendo programados e sendo alocados em cada maqueiro. A cada entrada de novo paciente no sistema

ocorre um novo sequenciamento, bem como a mudança de nível de prioridade dos pacientes. Se o tempo de transformação de prioridade for alcançado o paciente recebe uma priorização diferente no novo sequenciamento, até que todos os pacientes sejam deslocados e a estrutura de SolucaoFinal completa. O fim desse processo acontece quando não existem novos pacientes entrando no sistema e todos os pacientes anteriormente programados para serem deslocados já tiverem sido transportados.

4.4 VALIDAÇÃO DO MODELO

Foi realizada a validação do modelo matemático com um exemplo contendo 7 pacientes (mais o *job 0* fictício inicial) e 2 maqueiros. Foi realizada uma análise a cada entrada de novo paciente. Os tempos para cada prioridade foram os seguintes (Tabela 1):

Tabela 1 – Tempos por Nível de Prioridade

Nível de Prioridade (w)	Tempo (min)
4	1
3	2
2	3
1	4

Fonte: A autora (2021).

Na Tabela 2 estão representados os níveis de prioridade (*w*) originais dos pacientes por meio das cores amarelo, marrom e verde em 2 tonalidades (escuro e claro). O *setup* em branco representa o tempo de deslocamento entre o destino de um paciente e a origem do paciente imediatamente seguinte.

Tabela 2 - Níveis de prioridade por cor

w	Cor	
4	Amarelo	
3	Marrom	
2	Verde	
1	Verde claro	
Setup	Branco	

Fonte: A autora (2021).

Como dados de entrada tem-se os tempos de transformação de prioridades para teste sendo 1, 2, 3 e 4 minutos os respectivos tempos para os níveis 4, 3, 2 e 1. A Tabela 3 a seguir informa o instante de liberação (r_j), tempo de deslocamento do paciente j (p_j) e o nível de prioridade na chegada do paciente ao sistema (w).

Tabela 3 - Dados de entrada do sistema

Job	r_j (min)	p_j (min)	w
0	0	0	4
1	0	7	4
2	0	8	3
3	0	5	2
4	0	3	1
5	4	3	4
6	7	2	3
7	12	5	1

Fonte: A autora (2021).

Também como dado de entrada tem-se a matriz simétrica com o tempo de deslocamento entre os pacientes (Figura 5).

Figura 5 - Matriz com tempo de deslocamento entre os locais dos pacientes (minutos)

	Pac0	Pac1	Pac2	Pac3	Pac4	Pac5	Pac6	Pac7
Pac0	0	0	0	0	0	0	0	0
Pac1	0	0	5	4	5	4	5	4
Pac2	0	5	0	1	2	3	1	2
Pac3	0	4	1	0	1	1	2	3
Pac4	0	5	2	1	0	1	1	1
Pac5	0	4	3	1	1	0	2	1
Pac6	0	5	1	2	1	2	0	1
Pac7	0	4	2	3	1	1	1	0

Fonte: A autora (2021).

Considerando um número de maqueiros (m) e de pacientes (n), no total, para serem atendidos em diversos horário do dia, tem-se a primeira programação. Nesse sequenciamento serão programados os pacientes que entraram no primeiro horário como poder ser visto no sequenciamento a seguir. Na primeira rodada, com o relógio = 0, tem-se:

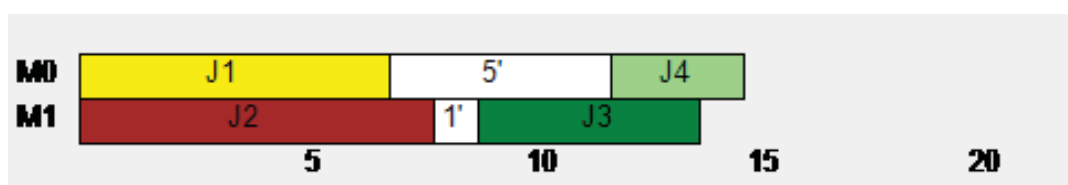
Figura 6 - Posição dos pacientes no vetor JobsParaConsiderar com $t = 0$

$t = 0$	JobsParaConsiderar:	0	1	2	3	4
		[0]	[1]	[2]	[3]	[4]

Fonte: A autora (2021).

O vetor JobsParaConsiderar, nesse instante, teve 4 pacientes (pacientes que tem seu instante de liberação menor ou igual ao instante do relógio) e a Figura 6 apresenta a configuração dos pacientes em cada posição do vetor. Tendo esses pacientes todos disponíveis para serem programados (pois o instante de liberação deles é 0), o modelo foi montado e posteriormente rodado. Como resultado o modelo gerou o sequenciamento a seguir da (Figura 7). Nesse caso não houve tempo ocioso dos maqueiros no sequenciamento pois o modelo encontrou a maneira mais eficiente programando pacientes imediatamente em seguida um do outro.

Figura 7 - Sequenciamento gerado pelo modelo com relógio = 0



Fonte: A autora (2021).

A partir disso se inicia o ressequenciamento propriamente dito, visto que a cada nova chamada o programa será rodado novamente. Alguns cenários são possíveis no ressequenciamento: pode haver pacientes que já estão em deslocamento, e como são consideradas máquinas paralelas sem preempção, o paciente não poderá sofrer qualquer alteração para outro maqueiro ou ter seu deslocamento interrompido. O maqueiro deve finalizar esse deslocamento e somente após isso poderá realizar uma futura chamada. Também pode haver pacientes que não tiveram seu deslocamento iniciado então eles podem sofrer qualquer mudança na posição de seu sequenciamento programado anteriormente. O próximo paciente recebe entrada no sistema com relógio igual a 4 (Figura 8).

Figura 8 - Posição dos pacientes no vetor *JobsParaConsiderar* com $t = 4$

$t = 4$	JobsParaConsiderar:	0	3	4	5
		[0]	[1]	[2]	[3]

Fonte: A autora (2021).

A partir desse segundo sequenciamento foi necessário utilizar uma tabela de tradução de *jobs* para a validação. Nesse momento da otimização havia 4 *jobs* na fila (0, 3, 4 e 5) e esses *jobs* ocuparam as posições 0, 1, 2 e 3. O *job* 0 permanece até o fim e a tradução foi feita da seguinte forma:

Figura 9 - Tabela de tradução para sequenciamento com relógio = 4

	J03		J05		J04	
Modelo	X011	C11	X030	C30	X320	C20
Traduzido	X031	C31	X050	C50	X540	C40
Valor	1	14	1	14	1	18
	E32=2		E54=5		E44=1	
					E43=3	
					E42=6	
					E41=10	

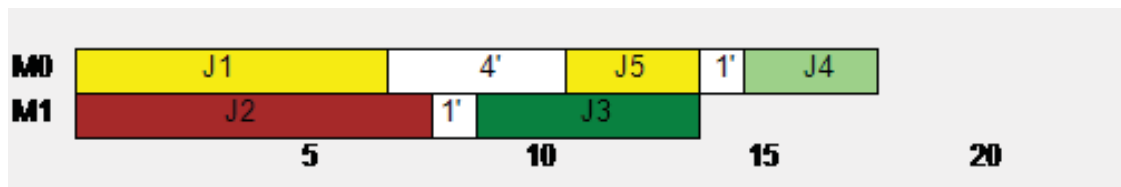
Fonte: A autora (2021).

Os valores ε_{jw} na Figura 9 dizem respeito ao atraso que irá ocorrer caso esse sequenciamento seja efetivado. O modelo já adianta o tempo de atraso, por exemplo, do *job* 4. A prioridade 1 tem o tempo de 4 minutos para ser atendido, após esse tempo o paciente sobe para a prioridade 2. Nesse caso a previsão é de que o paciente atrase 4 minutos em P1, 3 minutos em P2, 2 minutos em P3 e 1 minuto em P4. No *job* original é contabilizada a soma dos atrasos sendo 10 minutos no total (4 + 3 + 2 + 1). Na P2 aparece o valor de 6 pois se trata de 3 + 2 + 1 (minutos), assim como na P3 tem-se o valor 3 de 2 + 1 minutos e em P4 somente 1 minuto.

O sequenciamento para esse tempo do relógio é mostrado na Figura 10. No momento do ressequenciamento, como falado anteriormente, também pode haver pacientes que estão na fila mas ainda não foram deslocados. Esses pacientes serão incluídos na reprogramação podendo sofrer alteração no horário de início de deslocamento programado por causa do seu nível de prioridade. Um paciente com mais emergência pode entrar no sistema e após o ressequenciamento o mesmo será

atendido mais rapidamente. Foi o que aconteceu com o *job* 5 entrando nesse momento no sistema e após rodar o modelo o *job* 5 entrou na frente do *job* 4.

Figura 10 - Sequenciamento gerado pelo modelo com relógio = 4



Fonte: A autora (2021).

O instante em que o próximo paciente entrou no sistema foi em 7 minutos. Antes disso o sequenciamento programado continuou sendo processado. A Figura 11 a seguir apresenta a disposição dos pacientes no vetor *JosbParaConsiderar*.

Figura 11 - Posição dos pacientes no vetor *JosbParaConsiderar* com $t = 7$

t=7	JobsParaConsiderar:					
	0	3	4	5	6	
	[0]	[1]	[2]	[3]	[4]	

Fonte: A autora (2021).

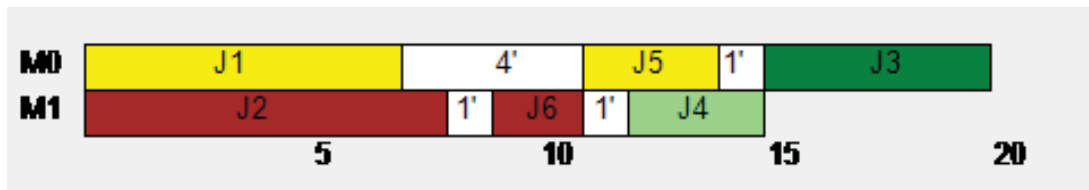
A tabela de tradução do sequenciamento está representada na Figura 12. O *job* 3 recebe uma prioridade mais urgente, porém, ao mesmo tempo, o *job* 6 entra no sistema com uma prioridade maior. O sequenciamento mais eficiente encontrado pelo modelo está representado na Figura 13.

Figura 12 - Tabela de tradução para sequenciamento com relógio = 7

	J05		J06		J03		J04	
Modelo	X030	C30	X041	C41	X310	C10	X421	C21
Traduzido	X050	C50	X061	C61	X530	C30	X641	C41
Valor	1	14	1	11	1	20	1	15
	E54=5		E34=6		E43=3			

Fonte: A autora (2021).

Figura 13 - Sequenciamento gerado pelo modelo com relógio = 7



Fonte: A autora (2021).

Chamada após chamada os pacientes são programados e tendo seus lugares na fila de deslocamento alterados conforme cada prioridade. A cada ressequenciamento o instante final de deslocamento da atividade fictícia inicial (C_{0k}) é atualizado para o maior valor entre o instante de término do último paciente de cada maqueiro ou o instante do relógio. O exemplo foi finalizado com relógio = 12 com o vetor JobsParaConsiderar (Figura 14) da seguinte forma:

Figura 14 - Posição dos pacientes no vetor JobsParaConsiderar com $t = 12$

t=12	JobsParaConsiderar:	0	3	7
		[0]	[1]	[2]

Fonte: A autora (2021).

A tabela de tradução é mostrada na Figura 15 e em seguida na Figura 16 o sequenciamento final visto que o job 7 é o último paciente do exemplo.

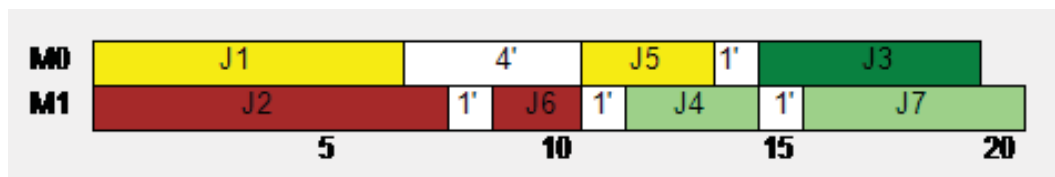
Figura 15 - Tabela de tradução para sequenciamento com relógio = 12

	J03		J07	
Modelo	X010	C10	X021	C21
Traduzido	X030	C30	X071	C71
Valor	1	20	1	21

E34=6

Fonte: A autora (2021).

Figura 16 - Sequenciamento gerado pelo modelo com relógio = 12



Fonte: A autora (2021).

Falando novamente sobre a transformação de prioridade: cada paciente, ao entrar no sistema, tem um *due date* associado ao seu nível de prioridade. Se caracteriza pelo instante de tempo máximo em que o paciente precisa ser entregue em seu destino (soma do instante de entrada no sistema, tempo de transformação de prioridade e o deslocamento propriamente dito). Caso o paciente não seja atendido dentro desse limite (*due date*), os minutos seguintes começam a ser contabilizados como atraso no nível de prioridade inicial. O paciente recebe um novo *due date* em que se soma o *due date* anterior com o tempo de transformação de prioridade do novo nível (após a transformação para prioridade mais urgente).

Sendo o paciente entregue em seu destino dentro desse novo limite, não há mais penalização de tempo. Caso ultrapasse esse limite, os minutos seguintes são novamente contabilizados como atraso em todos os níveis que o paciente passou, até que finalmente o paciente seja deslocado e entregue ou que novamente uma mudança de prioridade e um novo *due date* sejam gerados.

Após a demonstração de como é o mecanismo de atualização dos níveis de prioridade e ressequenciamento, será visto como foram desenvolvidos os testes computacionais para o estudo desses tempos de transformação dos níveis de prioridade.

5. TESTES COMPUTACIONAIS

Nesse capítulo são apresentados a descrição das instâncias geradas com valores baseados em dados reais e os resultados dos testes computacionais para estes problemas resolvidos pelo método exato implementado.

5.1 DESCRIÇÃO DAS INSTÂNCIAS

As instâncias utilizadas para teste foram extraídas do trabalho dos autores Ton e Da Silva (*in press*) sendo os conjuntos de instâncias de testes reais. Essas instâncias contêm as seguintes informações: número de pacientes a serem transportados chamado de *job* (n), número de maqueiros (8), número de quartos (30 origens e destinos), lista de níveis de prioridades (1, 2, 3 e 4), peso de cada prioridade na função objetivo (em potência de 2 sendo $2^0, 2^1, 2^2$ e 2^3 para os níveis 1, 2, 3 e 4 respectivamente). Pode-se encontrar também 3 vetores de tamanho n contendo a prioridade inicial de cada *job*, a zona de origem e a zona de destino de cada paciente. No fim do arquivo encontra-se a matriz de distâncias entre os 30 quartos considerados no trabalho.

No total foram 108 instâncias com combinações das seguintes informações: números de *jobs*, cenários e repetições de cada combinação em cada cenário. Os números de pacientes foram 20, 30 e 40. Os cenários foram 1, 2 e 3 e representam a prevalência dos níveis de prioridade nos pedidos gerados. De acordo com o artigo Ton e Da Silva (*in press*), em $SC = 1$ existem 10% dos pacientes iniciando com $w = 4$ (prioridade), tendo o resto dos pedidos distribuídos com uniformemente com 30% de probabilidade nos níveis 1, 2 e 3. No cenário 2 tem-se 25% dos pedidos em cada nível de prioridade. O terceiro cenário contém, em média, 34% de pacientes iniciando com prioridade nível 4, enquanto os outros níveis de prioridade de 22% dos pacientes cada.

Pode-se entender os 3 diferentes cenários como momentos diferentes na rotina de um hospital (manhã, tarde e noite), ou mesmo 3 setores diferentes de um hospital: urgência e emergência, setor de consultas e internação hospitalar. Desse modo veremos o comportamento desses tempos de transformação de prioridade para cada um dos diferentes ambientes fazendo a diferenciação na quantidade de pacientes com cada prioridade. A diversificação dos cenários ocorreu por conta da busca de aproximação do modelo à realidade e com isso analisar o comportamento

dos sequenciamentos, além de observar quais mudanças relevantes aconteceriam entre os sequenciamentos (makespan, atraso para iniciar o deslocamento, entre outros).

A repetição das instâncias ocorre 12 vezes e acontece o seguinte: tem-se um cenário e uma porcentagem de pacientes com cada prioridade. No momento da geração dessas instâncias (artigo citado) foram mantidas as características de cada paciente (tempo de deslocamento, instante de liberação, entre outras). O único elemento modificado foi o nível de prioridade. Então respeitando a porcentagem de cada cenários, os níveis de prioridade foram trocados entre os pacientes 12 vezes, em cada uma das repetições. Isso gerou uma diversificação nas instâncias, aumentando as possibilidades de diferentes sequenciamentos.

Para melhor entendimento sobre as instâncias, será descrita a instância 20N1SC9REP: contém 20 pacientes, o cenário é o 1 o qual foi gerado tendo 10% dos pacientes iniciando com $w = 4$ (2 pacientes) e 30% dos pacientes para cada um dos outros níveis (6 pacientes em cada uma das prioridades 3, 2 e 1), totalizando 20 *jobs*. O 9 significa o número da repetição da instância como uma espécie de variação de Monte Carlo. Da repetição 1 até a 12 tem-se o número de pacientes, número de maqueiros, origem, destino, instante de liberação do *job* mantidos, modificando-se apenas o nível de prioridade de cada *job*, sendo mantidas as porcentagens de acordo com cada cenário. Essas repetições ampliam a variabilidade e abrangência dos testes causando ambientes diferentes a serem analisados.

Após a descrição das instâncias serão apresentados a seguir os tipos de testes realizados para ser realizada a comparação de desempenho que baseou esse estudo.

5.2 MODELO SEM TRANSFORMAÇÃO DE PRIORIDADE (TON E DA SILVA)

O modelo desenvolvido por Ton e Da Silva (*in press*) se trata do funcionamento atual do hospital Canadense o qual será utilizado como parâmetro para analisar o desempenho do modelo matemático desenvolvido. De acordo com o trabalho de Ton e Da Silva (*in press*), o hospital utiliza os níveis de prioridade sem a transformação (mesmo nível do início ao fim), tendo os tempos máximos esperados para atendimento de 60, 20, 7 e 0 minutos para as prioridades 1, 2, 3 e 4, respectivamente.

5.3 FERRAMENTA DE *DESIGN* DE EXPERIMENTOS (*IRACE*)

Tendo conhecimento do *baseline*, o próximo passo da pesquisa foi trabalhar e analisar os tempos de atendimento para encontrarmos os tempos de transformações que minimizassem o tempo de espera dos pacientes e o tempo ocioso de trabalho dos maqueiros. Para isso, foi escolhido usar uma ferramenta de *design* de experimentos chamada *irace*. Segundo Stützle et al. (2016), o *irace* recebe alguns parâmetros de entrada, instâncias e o modelo a ser utilizado. Com isso a ferramenta implementa procedimentos de configuração automática (pré-estabelecidas de acordo com o problema em questão) e tem como objetivo encontrar as melhores combinações de parâmetros de um otimizador. Esses testes foram realizados no *cluster* Graham da compute Canada, utilizando 4 processadores Intel E7 4850 v4 Broadwell de 2.1GHz e 4Gb de memória RAM.

O *irace* recebe como dado de entrada os indicadores que buscamos avaliar e um intervalo dos parâmetros para limitar a abrangência dos testes. O indicador utilizado foi o *makespan* (MKS) e foram dados 4 parâmetros: os tempos de transformação de cada um dos níveis de prioridade (P1, P2, P3 e P4). Esses intervalos foram baseados nos tempos originais do hospital canadense (60, 20, 7 e 0 minutos) e os próprios intervalos já garantiram uma restrição importante que é $P1 > P2 > P3 > P4$. Outro fator que influencia a confiabilidade é o tamanho do intervalo disponibilizado de cada parâmetro cujo IRACE realizará as combinações. Os intervalos foram [30,60], [15,25], [7,14] e [1,3] para os níveis 1, 2 3 e 4 respectivamente.

Foi configurado no *irace* o máximo de 20 mil rodadas e 108 instâncias. Quanto mais rodadas forem liberadas, maior a confiabilidade do resultado. O *irace* gera combinações dos 4 parâmetros dentro do intervalo e obtém um resultado como saída. A cada resultado de tempo de transformação encontrado o *irace* exclui as combinações com piores tempos. As combinações vão sendo testadas e no fim retornam-se as melhores. O *irace* retornou os 4 melhores resultados para os tempos de transformação de prioridades P1, P2, P3 e P4.

Os valores que se repetiram em 3 dentre as 4 melhores políticas foram 56, 16, 13 e 2 minutos para as prioridades 1, 2, 3 e 4 respectivamente. Esses números dizem respeito aos melhores tempos de espera por prioridade, resultando no tempo de transformação de prioridade que é um dos objetivos desse estudo. Caso um

paciente não seja deslocado dentro desse período, ele recebe uma prioridade mais urgente e começa a ser contado atraso para iniciar o deslocamento.

Tabela 4 - Resumo do uso do Irace

Indicador	Parâmetro	Intervalo (Inteiros)	Valor (min)
<i>Makespan</i> (MKS)	P1	[30,60]	56
	P2	[15,25]	16
	P3	[7,14]	13
	P4	[1,3]	2

Fonte: A autora (2021).

Esse processo todo foi realizado com o objetivo de melhorar o tempo de deslocamentos do hospital, bem como diminuir os atrasos no deslocamento dos pacientes e tempos ociosos dos maqueiros. A Tabela 4 nos mostra o indicador levado em consideração, os 4 parâmetros, os intervalos testados ([30,60], [15,25], [7,14] e [1,3]) e os valores encontrados. Foram utilizados números inteiros para limitar o espaço de busca. Na prioridade 4 o tempo de 0 minutos foi excluído do estudo visto que, um paciente em estado gravíssimo recebe um atendimento imediato de uma equipe de suporte à vida. Essa equipe de funcionários realiza diversos procedimentos como estancamento de sangue, instalação de aparelhos de respiração e/ou medidores de parâmetros, entre outros. Sendo assim, são equipes diferentes para tipos de atendimento diferentes, no caso, não incluso nesse estudo.

5.4 TESTES COMPUTACIONAIS

Na segunda parte dos testes foram utilizados os vetores de parâmetros obtidos no Irace e rodados em um processador Intel® Core™ i5 8250 U, 8 GB de RAM. O sistema operacional foi o Windows 10 Home, linguagem de programação C#, IDE Visual Studio 2019 versão 16.5.4 e utilizado o solver CPLEX v. 20.10, com número de threads igual à 1 e os outros parâmetros *default*.

Considerando os sobrenomes dos autores, os testes realizados foram 3: modelo sem transformação de prioridade com os tempos atuais do hospital (TON e DA SILVA), modelo exato da dissertação (com transformação) com os tempos do hospital (PAZ) e por último o modelo exato da dissertação com os tempos encontrados

a partir do uso do *irace* (PAZ+IRACE). Os testes foram realizados dessa maneira para ser possível a comparação entre o artigo citado e as possíveis melhoras propostas nesse trabalho, incluindo a utilização do resultado do *irace*. Foram gerados novos resultados com as seguintes configurações (Tabela 5):

Tabela 5 - Testes realizados

Referência	Modelo	Tempo de Transformação (T _{jw})
TON E DA SILVA	Sem <i>upgrade</i>	Hospital (60,20,7,0)
PAZ	Modelo Exato Dissertação (com <i>upgrade</i>)	Hospital (60,20,7,0)
PAZ + <i>Irace</i>	Modelo Exato Dissertação (com <i>upgrade</i>)	<i>Irace</i> (56,16,13,2)

Fonte: A autora (2021).

Tendo ciência dos tipos de testes realizados, serão apresentados no próximo capítulo os resultados dos testes computacionais, bem como a análise e comparações pertinentes ao estudo.

6. RESULTADOS COMPUTACIONAIS

Para a análise foram verificados os intervalos de confiança 95% utilizando T de *Student* para cada parâmetro considerado (*makespan*, atraso, tempo de resposta, resposta por prioridade, ociosidade dos maqueiros) e consideradas as 3 configurações anteriormente citadas na Tabela 5.

6.1 MAKESPAN

A primeira análise foi feita considerando o *makespan* dessas instâncias. O estudo considerou o início do relógio às 8 horas da manhã (28.800 segundos), a Tabela 6 apresenta a média do *makespan* nos 3 cenários de testes e as diferentes quantidades de pacientes.

Tabela 6 - Média de Makespan em segundos

SC	N	MÉDIA DE MKS (s)		
		TON E DA SILVA	PAZ	PAZ + IRACE
1	20	30861,00	30864,58	30864,58
	30	31543,67	31545,58	31551,83
	40	32962,83	32944,42	33008,92
2	20	31227,33	31201,83	31199,00
	30	32176,25	32140,42	32119,08
	40	32913,58	32883,58	32868,50
3	20	31280,08	31241,08	31212,50
	30	31992,58	31983,92	32030,00
	40	33215,25	33216,17	33163,36

Fonte: A autora (2021).

Visto que os valores de *makespan* apresentados foram muito semelhantes, não serão apresentados os intervalos de confiança pois a maior margem de erro (valor que deve ser subtraído e somado a média originando o intervalo de confiança) quando calculado foi de 0,003%.

É possível observar no cenário 1 que para 20 pacientes os valores de *makespan* foram praticamente os mesmos, comparando-se os 3 experimentos não houve melhora e nem piora no tempo de entrega do último paciente. O mesmo acontece com 30 e 40 pacientes. De acordo com o artigo de Ton e Da Silva (*in press*), em $SC = 1$ existem 10% dos pacientes iniciando com $w = 4$ (prioridade), tendo o resto

dos pedidos qualquer outro nível de serviço com probabilidade uniforme (30% pacientes nos níveis 3, 2 e 1). A partir da realização dos testes foi percebido que a pequena quantidade de *jobs* tem influência na variação do sequenciamento, bem como a quantidade de *jobs* em cada prioridade. Veremos a frente que nesse nível de prioridade 4 é que são causados os maiores atrasos pois o tempo para deslocamento é justo. Como nesse cenário a minoria dos *jobs* tem o nível de prioridade $w = 4$ no início, menor será o atraso e consequente menor *makespan*.

Nos cenários 2 e 3 os valores se mantêm bem próximos novamente. Sendo assim, é possível concluir que o *makespan* é um fator que deve ser levado em consideração e nesse caso o que pode diminuir esse valor é o tempo de *setup* entre os pacientes. O baixo número de pacientes contribui para a pouca variação desse indicador e percebe-se que ele sozinho não é suficiente para analisarmos o desempenho de cada tipo de teste, sendo necessário avaliar outras variáveis a seguir.

6.2 ANÁLISE DA FUNÇÃO OBJETIVO (FO)

A função objetivo foi analisada considerando o *makespan*, descrito na seção 6.2, somado aos atrasos em cada prioridade. O peso na função objetivo dos atrasos se dá em potência de 2, sendo 2^0 , 2^1 , 2^2 e 2^3 para os níveis 1, 2, 3 e 4 respectivamente. A Tabela 7 apresenta a média e o intervalo de confiança 95% utilizando t de *Student* visto que a amostra foi somente 12 instâncias para cada cenário e número de pacientes.

Tabela 7 - Função Objetivo em segundos

SC e N	FUNÇÃO OBJETIVO (s)					
	TON E DA SILVA		PAZ		PAZ + IRACE	
	Média	Int. Conf.	Média	Int. Conf.	Média	Int. Conf.
1						
20	30861,00	[30840,09; 30881,91]	30864,58	[30833,10; 30896,07]	30864,58	[30833,10; 30896,07]
30	31543,67	[31509,65; 31577,68]	31545,58	[31511,05; 31580,12]	31551,83	[31521,30; 31582,37]
40	33301,50	[32589,07; 34013,93]	33827,75	[32432,33; 35223,17]	33562,25	[32834,67; 34289,83]
2						
20	31269,33	[31241,66; 31297,00]	31289,17	[31118,46; 31459,88]	31199,00	[31199,00; 31199,00]
30	32820,25	[32095,64; 33544,86]	33149,75	[32142,08; 34157,42]	32516,42	[31943,44; 33089,39]
40	33497,58	[32602,19; 34392,98]	32883,58	[32771,51; 32995,66]	33121,17	[32588,73; 33653,60]
3						
20	31984,75	[30547,73; 33421,77]	31295,08	[31218,72; 31371,45]	31212,50	[31147,14; 31277,86]
30	33288,58	[32140,82; 34436,35]	33199,25	[32063,62; 34334,88]	32282,00	[31740,80; 32823,20]
40	34204,58	[33133,81; 35275,35]	35063,50	[33131,55; 36995,45]	33664,83	[32520,00; 34809,67]

Fonte: A autora (2021).

Como explicado anteriormente, na função objetivo existem 2 termos: *makespan* e o somatório dos atrasos de cada *job* em cada prioridade. Observando-se a Tabela 4 é visível que os intervalos de confiança possuem uma área de interseção, os resultados das FO são bem próximos, tendo como unidade segundos. Os coeficientes dos atrasos por prioridade exercem um trabalho considerável na FO quando consideramos os diferentes cenários, bem como a quantidade de *jobs*.

No cenário 1 apenas 10% dos *jobs* tem prioridade nível 4 e com 20 e 30 *jobs* não houve atrasos, igualando-se a média da FO a média do *makespan* (Tabela anterior). Tendo ciência de que o número de maqueiros foi mantido do início ao fim dos testes (8 maqueiros), para 40 *jobs* o hospital apresenta o menor *makespan* com alguns pacientes atrasados. O teste TON E DA SILVA apresentou uma média geral de *makespan* menor do que os testes PAZ e PAZ + IRACE, porém estatisticamente estão dentro do mesmo intervalo de confiança.

No segundo cenário, em que há um equilíbrio na quantidade de *jobs* sendo 25% em cada prioridade, pode-se perceber uma média geral inferior do teste PAZ + IRACE em relação aos testes TON E DA SILVA e PAZ em 20 e 30 *jobs*, porém estatisticamente estão dentro do mesmo intervalo de confiança. No cenário 3, mesmo havendo uma maioria de *jobs* com prioridade 4, onde o atraso pesa mais na função objetivo, o teste PAZ + IRACE apresentou um melhor desempenho na otimização e menos atrasos com até 1.398,67 segundos de diferença do teste PAZ com 40 pacientes. O fato de haver uma mudança de 0 para 2 minutos diminui o tempo de atraso, ter mais *jobs* com prioridade 4 faz com que sejam atendidos mais rapidamente e tendo menos atraso haverá uma menor contabilização na função objetivo mesmo com peso de 2^3 .

Após a apresentação desses dados de função objetivo pode-se entender que o teste PAZ + IRACE apresenta um desempenho um pouco melhor em relação aos outros 2 testes conforme aumenta-se a quantidade de *jobs* porém estão dentro do mesmo intervalo. O fato de todos estarem no mesmo intervalo de confiança no indica algumas coisas: que é possível fazer uma mudança de prioridades sem alterar o *makespan*, sem aumentar o tempo de espera e o tempo ocioso dos maqueiros, que é algo esperado. A modificação das prioridades não causa uma mudança no tempo dos deslocamentos em si, mas sim nos tempos de *setup* entre os pacientes visto que a sequência de atendimento será diferente.

6.3 TEMPO DE RESPOSTA POR PRIORIDADE

Esse campo extraído como resposta da otimização diz respeito à diferença do instante que o chamado chega no sistema até o instante que o maqueiro chega no paciente para iniciar o deslocamento. Um aspecto dessa análise é justamente a percepção de tempo de espera de cada paciente para ser deslocado, por isso observou-se a necessidade de fazer essa análise do tempo de resposta por prioridade. Se trata de um indicador de performance que não está atrelado a função objetivo, e é algo que pode ser mensurado na percepção dos pacientes. A Tabela 8 mostra como se deu os tempos de resposta na prioridade 1.

Tabela 8 - Análise do tempo de resposta médio de P1 (minutos)

SC e N	TEMPO DE RESPOSTA MÉDIO P1 (min)		
	TON E DA SILVA	PAZ	PAZ + IRACE
1			
20	3,43	3,69	3,72
30	4,60	4,92	4,96
40	13,89	13,93	13,58
2			
20	4,21	3,54	3,45
30	8,85	8,76	7,89
40	15,86	15,09	14,47
3			
20	5,48	5,54	5,10
30	10,57	11,28	11,67
40	16,14	16,68	16,31

Fonte: A autora (2021).

No cenário 1 os valores foram bem próximos tendo uma interseção nos seus intervalos de confiança. O tempo do teste TON E DA SILVA para essa prioridade é de 60 minutos e mesmo após a utilização do *irace* o valor de tempo de transformação de prioridade ficou em 56 minutos. Com o tempo de resposta variando entre 7 e 12 minutos chega-se a conclusão de que nessa prioridade não existem atrasos. Nos cenários 2 e 3 o teste PAZ + IRACE teve um melhor desempenho fazendo com que a espera do paciente fosse menor, mesmo sendo uma melhora de menos de 1 minuto.

Tabela 9 - Análise do tempo de resposta médio de P2 (minutos)

SC e N	TEMPO DE RESPOSTA MÉDIO P2 (min)		
	TON E DA SILVA	PAZ	PAZ + IRACE
1			
20	3,17	3,12	3,12
30	4,51	4,55	4,56
40	12,55	12,04	11,77
2			
20	2,88	2,60	2,61
30	7,33	7,85	7,17
40	12,66	12,31	10,88
3			
20	7,14	7,42	6,50
30	11,05	10,27	9,11
40	12,59	13,10	11,42

Fonte: A autora (2021).

Sabe-se que o tempo na prioridade 2 é um pouco mais justo sendo 20 minutos para os testes TON E DA SILVA e PAZ e 16 minutos para o PAZ + IRACE. Tendo 3 cenários e 3 diferentes números de pacientes, na Tabela 9 percebe-se que o teste PAZ + IRACE bateu os tempos do TON E DA SILVA e PAZ em 8 de 9 ambientes, mesmo não havendo diferença estatística. A maior diferença de tempo de resposta chegou a quase 2 minutos nessa prioridade 2. A seguir será vista a Tabela 10 com o atraso na prioridade 3.

Tabela 10 - Análise do tempo de resposta médio de P3 (minutos)

SC e N	TEMPO DE RESPOSTA MÉDIO P3 (min)		
	TON E DA SILVA	PAZ	PAZ + IRACE
1			
20	3,00	3,02	3,02
30	4,45	4,59	4,22
40	8,77	8,67	10,60
2			
20	4,03	3,45	3,46
30	6,24	5,95	7,17
40	7,98	7,57	8,77
3			
20	4,22	3,79	4,83
30	7,76	8,06	8,92
40	11,45	8,94	11,00

Fonte: A autora (2021).

Nessa prioridade os tempos de transformação eram 7 minutos para os testes TON E DA SILVA e PAZ e 13 minutos para PAZ + IRACE. Novamente em 9 ambientes (3 cenários e 3 números de pacientes) o teste PAZ + IRACE não teve melhor desempenho em 7. No cenário 1 tem-se valores em minutos bem próximos tendo uma diferença um pouco maior aumentando-se os pacientes para 40. No cenário 2 que é um ambiente equilibrado em relação ao número de pacientes com cada prioridade (25%) o teste PAZ teve melhor tempo de resposta. Ou seja, o modelo aqui desenvolvido, porém com o tempo original de espera do hospital. No terceiro cenário o teste PAZ + IRACE melhorou seu tempo aumentando-se o número de pacientes. Será apresentada a Tabela 11 com a análise do tempo da prioridade 4.

Tabela 11 - Análise do tempo de resposta médio de P4 (minutos)

SC e N	TEMPO DE RESPOSTA MÉDIO P4 (min)		
	TON E DA SILVA	PAZ	PAZ + IRACE
1			
20	1,85	2,45	2,45
30	3,38	3,10	3,34
40	9,63	7,19	9,08
2			
20	4,46	4,14	4,22
30	5,28	5,05	4,85
40	7,66	7,53	7,48
3			
20	4,67	4,70	4,74
30	6,79	7,07	7,54
40	9,10	9,20	8,70

Fonte: A autora (2021).

A prioridade de maior urgência (P4) tem os tempos para atendimento de 0 minutos nos testes TON E DA SILVA e PAZ, sendo 2 minutos no teste PAZ + IRACE. Na Tabela 8, no cenário 1 inteiro o teste PAZ + IRACE teve um tempo de resposta maior, mudando somente no cenário 2 com 30 *jobs*. Com 40 *jobs* os testes TON E DA SILVA E PAZ ficaram a menos de 1 minutos de diferença no tempo de resposta. Estatisticamente todos se encontram no mesmo intervalo de confiança, o que os coloca em situação de empate. No cenário 3 acontece o mesmo com o teste PAZ + IRACE tendo tempos piores e aumenta-se o número de pacientes obtém-se uma leve

melhora no tempo de resposta.

Observando os testes PAZ e PAZ+IRACE é visível que as médias de tempo variam de acordo com quantidade de P4 por cenário. No cenário 1 a quantidade é pequena de P4 no teste PAZ+IRACE, e a média de tempo de resposta é maior. À medida que essa quantidade vai aumentando, de acordo com os cenários, o teste PAZ tem seu desempenho alterado e o PAZ+IRACE garante o menor tempo. Pode-se perceber também que, aumentando o número de pacientes de prioridade 4, os tempos de resposta para atendimento vão diminuindo, consequência dos tempos mais curtos para iniciar o deslocamento e da urgência.

Não é necessário haver uma unanimidade visto que a diferença dos cenários pode dizer respeito a 3 hospitais diferentes (pronto-socorro e hospital das clínicas, por exemplo, tem uma dinâmica diferente de pacientes urgentes e para consulta), ou o mesmo hospital para diferentes horários do dia. No início da manhã pode-se ter maior número de pacientes se deslocando entre consultórios e exames, no horário do almoço aumentam-se os pacientes com alguma emergência causado por eventuais acidentes, e em um outro período do dia um equilíbrio recebendo todos os tipos de pacientes na mesma quantidade. De acordo com o perfil do hospital existe um conjunto de pesos específico para realização das atividades.

6.4 TEMPO OCIOSO (MAQUEIROS)

Nessa seção serão analisados os tempos ociosos identificados no sequenciamento. Todos os tempos que não forem de deslocamento e de *setup* entre os pacientes são considerados tempos ociosos, significando maqueiro parado. A Tabela 12 a seguir detalha um pouco dos cenários postos a teste.

Tabela 12 - Porcentagem de tempo ocioso dos maqueiros

SC	% OCIOSO					
	TON E DA SILVA		PAZ		PAZ + IRACE	
	Média	Int. Conf.	Média	Int. Conf.	Média	Int. Conf.
1	5,82%	[4.96, 6.68]	5,89%	[4.99, 6.79]	5,74%	[4.84, 6.64]
2	14,64%	[12.59, 16.70]	14,47%	[12.51, 16.44]	14,80%	[12.73, 16.87]
3	5,17%	[4.43, 5.92]	5,20%	[4.47, 5.92]	5,29%	[4.55, 6.04]

Fonte: A autora (2021).

Para essa análise não foi necessário considerar o número de *jobs* visto que a real diferença se encontra nos cenários com as prioridades sendo alteradas. Foi possível juntar as quantidades de *jobs* nos cenários totalizando 36 instâncias por cenário. No cenário 1 o teste PAZ + IRACE teve uma média um pouco menor e nos cenários 2 e 3 os tipos que tiveram o desempenho um pouco melhor foram o PAZ e o TON E DA SILVA, respectivamente.

Em todos os cenários os valores estão empatados estatisticamente sendo a diferença menor do que 1 minuto. Visto que em geral os tempos ociosos dos maqueiros são os mesmos, mudar o sequenciamento dos pacientes não interfere no tempo de trabalho e tempo ocioso dos maqueiros. Os tempos de deslocamento permanecem os mesmos, o que mudam são os tempos de *setup* entre os pacientes imediatamente seguidos. Nessa etapa é que pode haver um sequenciamento mais eficiente. Sabendo disso, pode-se concluir com essa tabela que é possível melhorar o sequenciamento sem que haja sobrecarga de trabalho para os maqueiros, sem aumentar o nível de serviço.

6.5 ATRASO POR PRIORIDADE

Esse tópico diz respeito a quantidade de *jobs* atrasados por prioridade. O número de *jobs* atrasados tem algumas variáveis como a diferença entre os cenários 1, 2 e 3, bem como o tempo de transformação de prioridade em si. As Tabelas 13 e 14 a seguir tem novamente o cenário (SC) e o número de pacientes (N) nas linhas; e nas colunas tem-se os tipos de ambiente e a quantidade de pacientes correspondente a cada prioridade. A linha cinza representa a porcentagem de pacientes com aquela prioridade em cada cenário.

No cenário 1 com 20 pacientes na prioridade 1 existem 30% dos pacientes com essa prioridade, totalizando 6 pacientes nessa prioridade. Mantendo-se o cenário, o número de pacientes se repete nas prioridades 2 e 3, mudando apenas para 10% dos pacientes (2) na prioridade 4, conforme a anterior explicação na seção 5.1 a respeito dos cenários.

Tabela 13 - Porcentagem de pacientes atrasados nas prioridades 1 e 2 (%)

SC e N	MÉDIA DE %ATRASSO P1				MÉDIA DE %ATRASSO P2			
	Quant.	TON E DA SILVA	PAZ	PAZ + IRACE	Quant.	TON E DA SILVA	PAZ	PAZ + IRACE
1	0,30				0,30			
20	6	0,00	0,00	0,00	6	0,00	0,00	0,00
30	9	0,00	0,00	0,00	9	0,00	0,00	0,00
40	13	0,00	0,00	0,00	13	0,16	0,11	0,19
2	0,25				0,25			
20	5	0,00	0,00	0,00	5	0,00	0,00	0,00
30	7	0,00	0,00	0,00	7	0,01	0,00	0,01
40	10	0,00	0,00	0,00	10	0,09	0,10	0,13
3	0,22				0,22			
20	4	0,00	0,00	0,00	4	0,00	0,00	0,04
30	6	0,00	0,00	0,00	6	0,13	0,06	0,03
40	8	0,00	0,00	0,00	8	0,15	0,24	0,16

Fonte: A autora (2021).

Na prioridade 1 não houve atrasos, isso se dá pelo alto tempo para a prioridade (60 e 56 minutos para os testes TON E DA SILVA /PAZ e PAZ + IRACE, respectivamente). Na prioridade 2 o teste PAZ + IRACE tem uma porcentagem maior de pacientes atrasados em todos os cenários, o que é compensado no cenário 3 com 30 pacientes, por exemplo, onde ele tem um desempenho bom de 2,78% de atraso em reação ao teste PAZ com 12,50% dos pacientes atrasados. O hospital no cenário 3 com 40 pacientes obteve um aumento significativo nos atrasos, sendo os tempos de transformação 20 minutos para os testes TON E DA SILVA /PAZ e 16 minutos para o teste PAZ + IRACE.

Na Tabela 11 a seguir encontra-se a porcentagem de pacientes atrasados nas prioridades 3 e 4. Verificando o atraso em P3 é visível que o teste PAZ + IRACE teve um desempenho significativamente melhor em todos os cenários e todas as quantidades de pacientes. Isso se dá porque o tempo dos testes TON E DA SILVA e PAZ é de 7 minutos enquanto do teste PAZ + IRACE é 13 minutos, logo esse tempo mais justo terá mais atrasos. Os atrasos na prioridade 4 acontecem em todos os 9 ambientes tendo o tempo mais curto ainda (0 e 2 minutos para testes os testes TON E DA SILVA /PAZ e PAZ + IRACE respectivamente). Em 8 de 9 ambientes o teste PAZ + IRACE teve um melhor desempenho, chegando a 16,67% de diferença no cenário 3 e 20 pacientes.

Tabela 14 - Porcentagem de pacientes atrasados nas prioridades 3 e 4 (%)

SC e N	MÉDIA DE %ATRASSO P3				MÉDIA DE %ATRASSO P4			
	Quant.	TON E DA SILVA	PAZ	PAZ + IRACE	Quant.	TON E DA SILVA	PAZ	PAZ + IRACE
1	0,30				0,10			
20	6,00	0,03	0,00	0,00	2,00	0,13	0,21	0,21
30	9,00	0,04	0,04	0,00	3,00	0,25	0,22	0,14
40	13,00	0,54	0,54	0,35	1,00	0,58	0,50	0,42
2	0,25				0,25			
20	5,00	0,00	0,00	0,00	5,00	0,23	0,13	0,08
30	7,00	0,19	0,17	0,01	9,00	0,46	0,45	0,26
40	10,00	0,40	0,33	0,16	10,00	0,66	0,64	0,57
3	0,22				0,34			
20	4,00	0,08	0,04	0,04	8,00	0,54	0,52	0,38
30	6,00	0,42	0,42	0,01	12,00	0,58	0,60	0,52
40	8,00	0,83	0,65	0,28	16,00	0,63	0,62	0,53

Fonte: A autora (2021).

Após analisar as 4 prioridades e os tipos em cada cenário é perceptível que a prioridade 4 tem as maiores diferenças de desempenho dos modelos, bem como com o tempo de transformação de prioridades. Levando em consideração que é a única prioridade que tem um tempo de 0 minutos para realizar o deslocamento, será analisado na próxima seção o tempo de atraso real nessa prioridade, bem como o comportamento desses atrasos ao modificarmos marginalmente o tempo de prioridade, se haverá ganho considerável para o deslocamento e principalmente para o tempo de espera do paciente.

6.6 TEMPO DE ATRASO REAL E ANÁLISE DA PRIORIDADE 4

Após verificar os atrasos por prioridade foi possível perceber que o nível de prioridade com maior índice de atrasos é o nível 4 (mais prioritário). O tempo justo é uma das justificativas para isso ocorrer visto que na prioridade 1, por exemplo, o tempo de transformação obtido no teste PAZ + IRACE foi de 56 minutos, sendo um tempo que permite uma folga para atendimento e atraso. Sabendo disso, foi feita uma análise mais detalhada dessa prioridade e como ela se comporta fazendo algumas alterações marginais nesse tempo de estudo em questão. Na Tabela 15 a seguir é possível ver que se tratando de tempo real de atraso na prioridade 4 o teste PAZ + IRACE teve uma média menor, independente do cenário. É o único teste com folga de 2 minutos para início do deslocamento então era algo esperado.

Tabela 15 - Atraso real da prioridade 4 em minutos

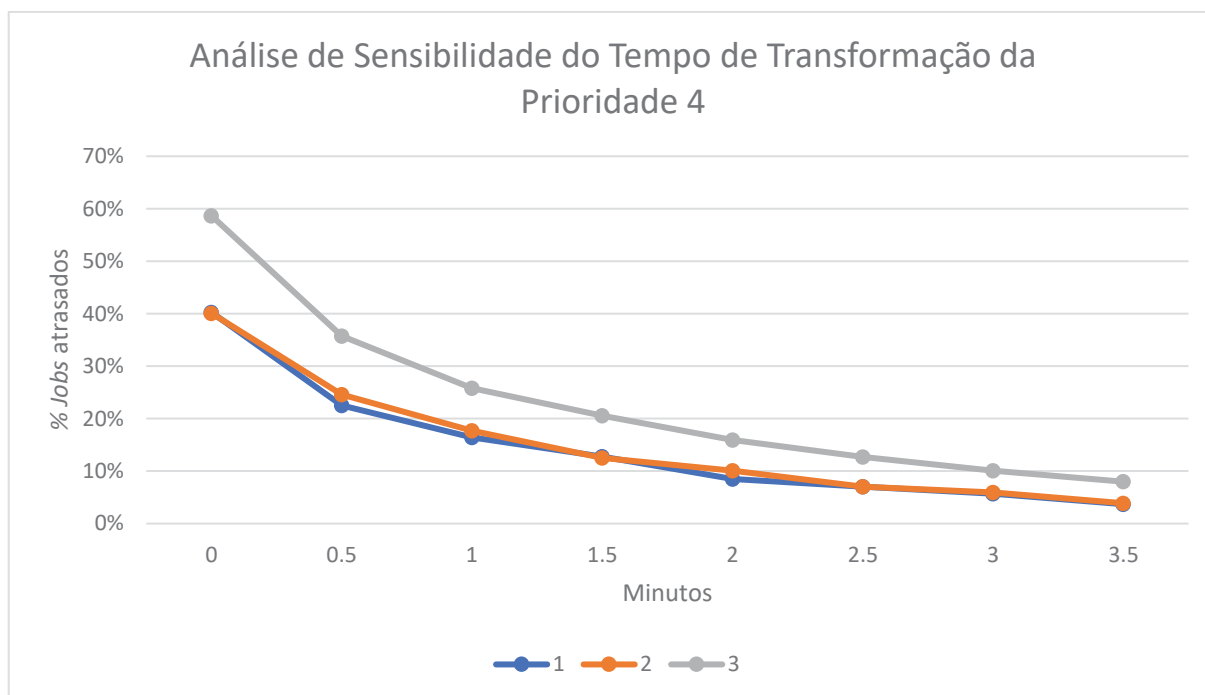
SC	ATRASSO REAL P4 (min)					
	TON E DA SILVA		PAZ		PAZ + IRACE	
	Média	Int. Conf.	Média	Int. Conf.	Média	Int. Conf.
1	2,53	[1,08; 3,97]	1,98	[0,73; 3,22]	1,49	[0,30; 2,68]
2	3,27	[2,49; 4,06]	3,35	[2,60; 4,10]	2,24	[1,63; 2,84]
3	4,95	[4,22; 5,69]	5,14	[4,39; 5,90]	3,81	[3,12; 4,49]

Fonte: A autora (2021).

Foram mantidas todas as características das instâncias exceto o tempo de transformação da prioridade 4. Para essa análise foi feito um gráfico (Figura 22) onde os eixos são:

- Eixo X: tempo de transformação da prioridade 4;
- Eixo Y: a média da porcentagem de *jobs* atrasados na prioridade 4 sobre o tempo de transformação de prioridade em questão + 1. Esse valor no eixo Y também pode ser entendido como a variação dos *jobs* atrasados ao se modificar 1 unidade do limite, verificando assim o ganho em atraso ao aumentar ou diminuir o tempo da prioridade em questão. Buscamos descrever as alterações causadas no atraso para uma mudança unitária no tempo de transformação.

Figura 17 - Gráfico para Análise de Sensibilidade do tempo de transformação da prioridade 4



Fonte: A autora (2021).

Algumas conclusões podem ser tomadas observando-se a Figura 22. Ao movimentar-se no eixo X do local 2 minutos, que foi o valor encontrado com o uso do *irace*, para a esquerda tem-se a diminuição dos tempos de transformação para a prioridade 4. Ao movimentar-se para a direita tem-se um aumento nesse tempo. Uma mesa análise pode ser feita no eixo Y em relação a porcentagem de pacientes atrasados nessa prioridade. As linhas do gráfico traduzem os 3 cenários estudados (1, 2 e 3) com as diferentes porcentagens e quantidades de pacientes com cada prioridade.

Analisando a curva do cenário 3, onde a maioria dos *jobs* tem prioridade 4, percebe-se que ao aumentar de 0 para 1 minuto o tempo na prioridade 4, tem-se uma queda de 60% de atraso no deslocamento para apenas 25% de pacientes atrasados. Nos cenários 2 e 3 também é perceptível a queda, porém, em menor escala (de 40% para aproximadamente 25% de pacientes atrasados).

Ao realizar a mudança de 3 minutos para 3,5 minutos a diferença na diminuição de pacientes atrasados é quase imperceptível. É visível que a dinâmica entre tempo de prioridade e pacientes atrasados não segue um padrão linear, é uma função exponencial. A partir de um certo valor de tempo de transformação no eixo X

a linha se mantém de certa forma constante, significando que a partir desse tempo não é válida a mudança aumentando o tempo de transformação pois não há melhora significativa na diminuição de pacientes atrasados nessa prioridade. O *irace* deu como resultado 2 minutos e a partir desse tempo o ganho em diminuição de atrasos é pequeno.

Deslocando-se no eixo X para a direita a partir do tempo de transformação igual a 0 é visível o ganho em atrasos só de realizar um aumento de 30 segundos a 1 minuto. Acontece uma queda na porcentagem de atrasos. O deslocamento em si pode demorar um tempo a mais para começar visto que o tempo do hospital atual é 0 minutos, porém esse tempo irá refletir mais no instante de término de deslocamento do paciente e menos no atraso da função objetivo. Esse é um fato importante pois a prioridade 4 tem coeficiente 2^3 na FO. O retorno será melhor na minimização da função objetivo diminuindo os atrasos mesmo que haja um aumento pequeno no instante de término no deslocamento do paciente.

Aumentar um pouco o tempo de transformação de prioridades aumenta o tempo de percepção de espera dos pacientes. Entretanto, um fator psicológico envolve o entendimento de ser atendido dentro do tempo estipulado ser melhor do que um tempo de 0 minutos para atendimento e a chance de o paciente ter seu deslocamento atrasado ser de 60%.

Em resumo, entende-se que tudo depende da política do hospital, o que é considerado como um nível de serviço eficiente, tanto relacionado ao tempo de espera do paciente, como nível de serviço dos maqueiros, por exemplo. Ou mesmo qual variação pode ser feita para diminuir a percepção do tempo de espera do paciente e que traga um benefício real para a rotina do hospital. A percepção de tempo de espera do paciente é um dos fatores mais importantes nesse estudo, bem como a otimização em si. É necessário pesar e equilibrar todas as variáveis estudadas como, por exemplo, a diminuição ociosidade dos maqueiros que tem como consequência maqueiros menos tempo parados e em deslocamento; o atraso e espera para atendimento por prioridade, o *makespan* em si, a função objetivo e por fim a análise mais detalhada da prioridade 4. Esse equilíbrio envolve até mesmo aumentar um pouco o tempo para transformação de prioridade em detrimento da função objetivo que busca a minimização do *makespan* e dos atrasos de maneira geral.

Estamos lidando nesse trabalho com nível de serviço e com transformação no

nível de prioridade. Sendo assim, é perceptível que não há uma solução ótima para esse caso, o que existem são pontos que podem ser melhorados buscando um nível de serviço razoável atendendo os mais prioritários rapidamente e não deixando as menores prioridades aguardarem tanto assim, e respeitando o cenário do hospital (quantidade e perfil de prioridade dos pacientes).

7. CONCLUSÃO

Com o objetivo de minimizar o *makespan* no deslocamento intra-hospitalar realizado por macas, foi desenvolvido um modelo matemático exato baseado no artigo de Ton e Da Silva (*in press*) levando-se em consideração o número de pacientes, número de maqueiros, os níveis de prioridades de cada paciente e os tempos de deslocamentos entre as zonas do hospital. A diferença entre os modelos foi a adição da mudança de prioridade para os pacientes que tivesse o tempo de transformação de prioridade ultrapassado. Com isso ocorre a penalização de atrasos por nível de prioridade e cada nível com um peso diferente na função objetivo. As inserções de novas chamadas e a realocação do sequenciamento de produção foram baseadas na política 1 no artigo citado visto que teve melhor desempenho como política de ressequenciamento.

Foram utilizados dados reais, considerando assim, dados relevantes para a realização da pesquisa. As 108 instâncias utilizadas para teste foram retiradas do artigo citado anteriormente para ser possível comparação no desempenho do modelo matemático. Foi utilizado o *irace* como ferramenta para design de experimentos e baseado nos tempos atuais do hospital, chegamos aos tempos de transformação de prioridade 56, 16, 13 e 2 minutos para as prioridades 1, 2, 3 e 4 minutos respectivamente. Foram 3 tipos de testes: TON E DA SILVA, PAZ e PAZ + IRACE, e com os resultados encontrados foi feita a análise deste trabalho.

Após rodar as 108 instâncias para os 3 tipos de testes considerando 20, 30 e 40 pacientes, cenários 1, 2, e 3, e 12 repetições trocando os níveis de prioridades dos pacientes, os resultados encontrados foram analisados em diversos aspectos. Em relação ao *makespan* não foi encontrada diferença estatística nesse tempo limite. Ter um sequenciamento eficiente pode não causar uma diminuição do instante de término de deslocamento do último paciente, porém também se percebe que é possível ter um sequenciamento mais eficiente sem haver aumento do *makespan*.

No aspecto função objetivo foi perceptível que a maior diferença se deu no cenário 3 entre o *baseline* que foi o teste TON E DA SILVA e o teste PAZ+IRACE. O cenário 3 tem em sua maioria pacientes com a prioridade 4, esse nível também tem a maior taxa de atrasos e mesmo assim o modelo aqui desenvolvido juntamente com o tempo resultado do *irace* teve o menor tempo de função objetivo.

Os aspectos tempo médio de resposta por prioridades e atraso por prioridade

deixam perceptível que o nível de prioridade 4 tem o menor tempo de resposta, bem como os maiores números de pacientes atrasados. Ao analisar o tempo ocioso dos maqueiros por cenário e para cada um dos tipos de teste foi visto que não há diferença nesse tempo sem deslocamento. Pode-se entender que é possível haver um melhor sequenciamento sem aumentar o nível trabalho desses profissionais.

Em último lugar foi realizada uma análise mais detalhada da prioridade 4 por entendermos que algumas alterações podem ser feitas e com isso possíveis melhoras ou pioras no número de pacientes atrasados, ou no tempo de transformação desse nível de prioridade. Em minutos, o tempo de atraso nessa prioridade foi menor no modelo desenvolvido aqui nesse trabalho em relação aos outros 2 testes. Para verificar quais melhoras ainda poderiam ocorrer, foi feita uma análise de sensibilidade.

Foi apresentado um gráfico contendo no eixo X o tempo de transformação de prioridade e no eixo Y a média da porcentagem de *jobs* atrasados na prioridade 4 sobre o tempo de transformação de prioridade em questão + 1, bem como a variação dos *jobs* atrasados ao se modificar 1 unidade do limite, verificando assim o ganho em atraso ao aumentar ou diminuir o tempo da prioridade em questão.

Buscou-se descrever as alterações causadas no atraso para uma mudança unitária no tempo de transformação e como resultado foi percebido que a se trata de uma função exponencial. Ao mudar de 0 minutos (tempo atual do hospital) para 1 minuto ou 2 minutos (resultado do *irace*) ocorre uma diminuição significativa na porcentagem de pacientes atrasados. Tendo ciência de que a percepção de espera do paciente é um fator importante, também se chegou à conclusão de que um pequeno aumento no instante de término de deslocamento do paciente impacta menos a função objetivo do que um atraso na prioridade 4 que tem como coeficiente o peso 2^3 .

Como falado anteriormente, os 3 cenários retratados nos testes podem significar diferentes horários de atendimento do hospital, então para avaliar o desempenho de cada tipo de teste é preciso saber qual o nível de serviço o hospital está priorizando a cada momento. Estamos lidando nesse trabalho com nível de serviço e com transformação no nível de prioridade. É perceptível que não há uma solução ótima para esse caso, e que diversos são os pontos de a serem melhorados buscando um nível de serviço razoável atendendo os mais prioritários rapidamente e não deixando as menores prioridades aguardarem tanto assim, e respeitando o

cenário do hospital (quantidade e perfil de prioridade dos pacientes).

O modelo matemático aqui desenvolvido apresentou melhoras em alguns aspectos consideráveis como um tempo mais adequado nas prioridades mais urgentes, bem como um sequenciamento mais eficaz sem aumentar o nível de serviço dos funcionários. O principal fator a ser considerado é a diminuição nos atrasos nos pacientes realizando um pequeno aumento no tempo utilizado atualmente no hospital. Como proposta futura, aumentando e diversificando os cenários haverá outros detalhes e aspectos a serem analisados, bem como aumentando o número de pacientes para um número mais próximo da realidade entre 800 e 1000 pacientes. Para a resolução do modelo com esse tamanho de *jobs* é necessária a aplicação de heurísticas para tornar possível e viável o tempo de resposta computacional do estudo.

REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, G. A queuing model for health centre. **ICEIE 2010 - 2010 International Conference on Electronics and Information Engineering, Proceedings**, v. 1, n. Iceie, p. 542–547, 2010.

ARAGON, L. G.; CURE, L.; TIONG, E.; BUSH, R. Modeling and analysis of short-term work planning in inpatient care settings. **Operations Research for Health Care**, v. 19, p. 14–25, 2018. Elsevier B.V.

BEAUDRY, A.; LAPORTE, G.; MELO, T.; NICKEL, S. **Dynamic transportation of patients in hospitals**. 2010.

BHATTACHARJEE, P.; RAY, P. K. **Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections**. Elsevier Ltd, 2014.

BONTE, P.; ONGENAE, F.; DE TURCK, F. Towards Optimizing Hospital Patient Transports by Automatically Identifying Interpretable Causes of Delays. **International Journal of Software Engineering and Knowledge Engineering**, v. 29, n. 6, p. 819–847, 2019.

CHENG, T. C. E.; SIN, C. C. S. A state-of-the-art review of parallel-machine scheduling research. **European Journal of Operational Research**, v. 47, n. 3, p. 271–292, 1990.

CHENG, T. C. E.; DING, Q. The time dependent machine makespan problem is strongly NP-complete. **Computers and Operations Research**, v. 26, n. 8, p. 749–754, 1999.

CILDOZ, M.; IBARRA, A.; MALLOR, F. Accumulating priority queues versus pure priority queues for managing patients in emergency departments. **Operations Research for Health Care**, v. 23, p. 100224, 2019. Elsevier Ltd. Disponível em: <<https://doi.org/10.1016/j.orhc.2019.100224>>. .

CORDEAU, J. A Branch-and-Cut Algorithm for the Dial-a-Ride Problem A Branch-and-Cut Algorithm for the Dial-a-Ride Problem. , , n. October 2014, 2006.

CORDEAU, J. F.; LAPORTE, G. A tabu search heuristic for the static multi-vehicle dial-a-ride problem. **Transportation Research Part B: Methodological**, v. 37, n. 6, p. 579–594, 2003.

CORDEAU, J. F.; LAPORTE, G. The dial-a-ride problem: Models and algorithms. **Annals of Operations Research**, v. 153, n. 1, p. 29–46, 2007.

CORRÊA, H. L., GIANESI, I. G. N., CAON, M.. **Planejamento , Programação e Controle da Produção**. São Paulo: Atlas, 2006.

TON, V. M., DA SILVA, N., C. O.. **Real-Time Management of Patients' Transport Requests: An Empirical Study**. (2020). Artigo não publicado.

H. DERSHIN. QUALITY IMPROVEMENT FOR A HOSPITAL PATIENT TRANSPORTATION SYSTEM. , 1993.

FERRAND, Y. B.; MAGAZINE, M. J.; RAO, U. S.; GLASS, T. F. Managing responsiveness in the emergency department: Comparing dynamic priority queue with fast track. **Journal of Operations Management**, v. 58–59, n. March, p. 15–26, 2018. Elsevier. Disponível em: <<https://doi.org/10.1016/j.jom.2018.03.001>>. .

FRÖHLICH VON ELMBACH, A.; SCHOLL, A.; WALTER, R. Minimizing the maximal ergonomic burden in intra-hospital patient transportation. **European Journal of Operational Research**, v. 276, n. 3, p. 840–854, 2019.

FUCHIGAMI, H. Y.. **Introdução ao Sequenciamento da Produção**. Catalão: UFG, 2013.

GERCHAK, Y.; GUPTA, D.; HENIG, M. Reservation planning for elective surgery under uncertain demand for emergency surgery. **Management Science**, v. 42, n. 3, p. 321–334, 1996.

GRAHAM, R. L.; LAWLER, E. L.; LENSTRA, J. K.; KAN, A. H. G. R. Optimization and Approximation in Deterministic Sequencing and Scheduling: a Survey. **Discrete**

Optimization II, v. 5, p.287–326, 1979.

HAGEN, M. S.; JOPLING, J. K.; BUCHMAN, T. G.; LEE, E. K. Priority queuing models for hospital intensive care units and impacts to severe case patients. **AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium**, v. 2013, p. 841–850, 2013.

HANNE, T.; MELO, T.; NICKEL, S. Bringing robustness to patient flow management through optimized patient transports in hospitals. **Interfaces**, v. 39, n. 3, p. 241–255, 2009.

JACOBSON, S. H.; HALL, S. N.; SWISHER, J. R. **Discrete-Event Simulation of Health Care Systems**. 2013.

JOSEPH ADAMS, EGON BALAS, D. Z. The Shifting Bottleneck Procedure for Job Shop Scheduling. , 1998.

JI, M.; ZHANG, W.; LIAO, L.; CHENG, T. C. E.; TAN, Y. Multitasking parallel-machine scheduling with machine-dependent slack due-window assignment. **International Journal of Production Research**, v. 57, n. 6, p. 1667–1684, 2019. Taylor & Francis. Disponível em: <<https://doi.org/10.1080/00207543.2018.1497312>>. .

JIN, S.; JEONG, S.; KIM, J.; KIM, K. A logistics model for the transport of disaster victims with various injuries and survival probabilities. **Annals of Operations Research**, v. 230, n. 1, p. 17–33, 2015.

KOFMAN, E.; LIPPMAN, S. A. M/M/1 Dynamic Priority Queue With Optional Promotion. **Operations Research**, v. 29, n. 1, p. 174–188, 1981.

LAMIRI, M.; XIE, X.; DOLGUI, A.; GRIMAUD, F. A stochastic model for operating room planning with elective and emergency demand for surgery. **European Journal of Operational Research**, v. 185, n. 3, p. 1026–1037, 2008.

RICHARD LEWIS MATTSON, J. M. M. Dynamic queue prioritization by modifying priority value based on queue's level and serving less than a maximum number of requests per queue. , v. 12, n. 19, p. 2–4, 1999.

MAJZOUBI, F., BAI, L. H., HERAGU, S. S.. The EMS vehicle patient transportation problem during a demand surge. **JOURNAL OF GLOBAL OPTIMIZATION**. Vol: 79, Ed. 4, pg 989-1006, 2021. DOI: 10.1007/s10898-020-00965-1

MOLENBRUCH, Y.; BRAEKERS, K.; CARIS, A.; VANDEN BERGHE, G. Multi-directional local search for a bi-objective dial-a-ride problem in patient transportation. **Computers and Operations Research**, v. 77, p. 58–71, 2017. Elsevier. Disponível em: <<http://dx.doi.org/10.1016/j.cor.2016.07.020>>. .

PAJOUH, F. M.; KAMATH, M. Applications of Queueing Models in Hospitals. **MWAIS 2010 Proceedings**, p. 1–6, 2010. Disponível em: <<http://aisel.aisnet.org/mwais2010/23>>. .

PIMENTA, L. B. **Otimização no sequenciamento de produção em uma fábrica de materiais médico-hospitalares**. Juiz de Fora, p. 48. Trabalho de Conclusão de Curso – Universidade Federal de Juiz de Fora, 2008. Disponível em: http://www.ufjf.br/engenhariadeproducao/files/2014/09/2008_3_Lucas.pdf.

PINEDO, M.. **Scheduling - Theory, Algorithms, and Systems**. 5. ed. New Jersey: Prentice-hall, 2016.

PRADO, D.. **Teoria das filas e da Simulação**. 5ª ed. São Paulo, INDG, 2009.

SAMAH, S.; ARMEL, W. S.; STARKS, D. W. The use of simulation to reduce the length of stay in an emergency department. **Winter Simulation Conference Proceedings**, v. 2, p. 1907–1911, 2003.

SCIOMACHEN, A.; TANFANI, E.; TESTI, A. Simulation models for optimal schedules of operating theatres. **International Journal of Simulation: Systems, Science and Technology**, v. 6, n. 12–13, p. 26–34, 2005.

SEGEV, D.; LEVI, R.; DUNN, P. F.; SANDBERG, W. S. Modeling the impact of changing patient transportation systems on peri-operative process performance in a large hospital: Insights from a computer simulation study. **Health Care Management Science**, v. 15, n. 2, p. 155–169, 2012.

SIDDHARTHAN, K.; JONES, W. J.; JOHNSON, J. A.; et al. A priority queuing model to reduce waiting times in emergency care - Business - ProQuest. , 2010. Disponível em: <<http://search.proquest.com/business/docview/229695571/1393D67FC893FDDB7B6/11?accountid=38609>>. .

SPIEKSMAN, F. C. R.; WOEGINGER, G. J.; YU, Z. Scheduling with safety distances. **Annals of Operations Research**, v. 57, n. 1, p. 251–264, 1995.

TAN, K. W.; LAU, H. C.; LEE, F. C. Y. Improving patient length-of-stay in emergency department through dynamic queue management. **Proceedings of the 2013 Winter Simulation Conference - Simulation: Making Decisions in a Complex World, WSC 2013**, p. 2362–2373, 2013.

TAVAKKOLI-MOGHADDAM, R.; TAHERI, F.; BAZZAZI, M.; IZADI, M.; SASSANI, F. Design of a genetic algorithm for bi-objective unrelated parallel machines scheduling with sequence-dependent setup times and precedence constraints. **Computers and Operations Research**, v. 36, n. 12, p. 3224–3230, 2009.

VAN ESSEN, J. T.; HANS, E. W.; HURINK, J. L.; OVERSBERG, A. Minimizing the waiting time for emergency surgery. **Operations Research for Health Care**, v. 1, n. 2–3, p. 34–44, 2012.

LIN, S. W.; YING, K. C. Uniform Parallel-Machine Scheduling for Minimizing Total Resource Consumption with a Bounded Makespan. **IEEE Access**, v. 5, p. 15791–15799, 2017.

